

© 2012 Kyle M. Broom

EXPLICATION, SIMILARITY, AND ANALOGY:
A DEFENSE AND APPLICATION OF PHILOSOPHICAL METHOD

BY

KYLE BROOM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Philosophy
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Professor Gary Ebbs, Chair
Professor Robert Cummins
Professor Timothy G. McCarthy
Professor Steven J. Wagner

ABSTRACT

With his 1951 publication of “Two Dogmas of Empiricism”, W.V.O. Quine launched a series of arguments against the idea that analyticity – “truth in virtue of meaning alone” – could be a philosophically explanatory notion. While his rejection represents a significant philosophical stride in its own right, to which many in the contemporary philosophical scene pay verbal respects, the revolutionary consequences of this insight often go ignored today. Much of current professional philosophy in virtually every sub-discipline carries on as though analyticity were a viable notion, because much of it aims at conceptual analysis – the “discovery” of the meanings of philosophical concepts, such as ‘mind’, ‘truth’, ‘meaning’, ‘person’, and ‘right’. Rejecting analyticity as a viable philosophical notion undermines such efforts insofar as they aim at conceptual analysis.

If philosophers give up the notion of analyticity, as they should, they are left with the questions of what goals and methods are viable for the analysis of concepts. One option is to concentrate effort on *explication* of familiar concepts; explication is an alternative to conceptual analysis in that it does not aim to dig out analytic truths about given concepts. Rather, it identifies what is interesting about a certain vague concept, and generates a new, more rigorous and precise concept to replace the old one. The justification for the replacement, in the most appealing cases, is in the possibilities for understanding that are introduced by the new explication. Examples of explication in my sense are the replacement of ordinary grammatical terms by the operators of first order logic, replacement of ‘game’ with a decision theoretic notion of games in game theory, and replacement of ‘true’ with ‘true-in-L’ in Tarskian

semantics. Advances of this sort depart consciously in some ways from ordinary usage of the terms they focus on, but in so departing, they develop new possibilities and perspectives for rigorously understanding the world.

In the dissertation that follows, I not only recommend explication as a method, I also put it into practice. I develop an explication of the sameness and difference relations that relies solely on the relation of denotation between a predicate and objects of which it is true. Two things are the same with respect to R just in case there is at least one R-predicate that denotes them in common.

I used this explication to develop related explications of similarity and analogy. These latter explications, I use to critique the current fashion of research on similarity and analogy in cognitive psychology. These studies have produced an immense body of interesting and promising work in the past several decades; however, much of the research design and data interpretation harbors philosophical errors and so often imports unjustified prejudices from the researchers themselves. My explication offers alternative formulations of the most salient aspects of these research traditions, formulations that avoid the errors of past research and, more importantly, suggest new research questions and possibilities that would not occur under the previous, philosophically muddled paradigm.

The project closes with an application of my explication of analogy to a discussion and critique of the philosophical literature on analogical arguments. While philosophers and logicians have developed workable theories for deductive and inductive reasoning, very little work of promise has been done on analogical arguments. My explication of analogy is used to diagnose the problems of past

attempts by philosophers to develop logics for analogical arguments and suggest, in light of the psychological research on analogy and the use of analogies in scientific thought, that philosophers should focus on analogical arguments as creative, heuristic devices, rather than as truth-preserving inference structures.

*For Ellen and Bill, who gave me a life;
and Alexandra, who brought me back to it*

ACKNOWLEDGMENTS

This project developed steadily and then remained near completion but unfinished for several years. For both the initial well-supervised and nurtured progress as well as the final push to completion, I owe a huge debt of gratitude to my advisor, Gary Ebbs as well as to Steven Wagner. I also wish to thank my other committee members, Robert Cummins and Timothy McCarthy for their ready willingness to help me see the project through after the passage of so long a hiatus. Much of what I have managed to accomplish in the pages that follow has also been made possible by the outstanding graduate students who were active in the department during my time there; the atmosphere of collegiality and mutual philosophical interest gave our work a lived, communal quality I have rarely observed in other graduate programs. I especially wish to thank Daniel Estrada, Todd Kukla, Nathalie Morasch, Brandon Polite and David Rowland. Thanks as well to my partner, Alex, and to my parents, who have encouraged and supported me in this, as in all my endeavors.

TABLE OF CONTENTS

CHAPTER 1: REJECTING ANALYTICITY, REJECTING ANALYSIS	1
CHAPTER 2: EXPLICATION	63
CHAPTER 3: SAMENESS, SIMILARITY, AND COGNITIVE PSYCHOLOGY	95
CHAPTER 4: ANALOGY AND PSYCHOLOGY	143
CHAPTER 5: ANALOGIES IN PHILOSOPHY AND UNCHARTED SPACES: RESPECTIVISM AND STRUCTURE-PREDICATES	194
BIBLIOGRAPHY	238

CHAPTER 1

REJECTING ANALYTICITY, REJECTING ANALYSIS

I. *Prolog im Himmel*

Many philosophers understand themselves to be engaged in something one might (and many of them do) call “conceptual analysis”. This enterprise (were it possible) would involve figuring out the truth about what heady terms, such as ‘mind’, ‘good’, ‘meaning’, ‘fact’, and ‘truth’, refer to.¹ Paradigmatically, such solutions would follow from (or themselves be) analytic truths; thus: conceptual *analysis*. G. E. Moore provides a precise characterization of analysis in this sense (Moore 1942: 663):

If you are to “give an analysis” of a given *concept* which is the *analysandum*, you must mention, as your *analysans*, a *concept* such that (a) nobody can know that the *analysandum* applies to an object without knowing that the *analysans* applies to it, (b) nobody can verify that the *analysandum* applies without verifying that the *analysans* applies, (c) any expression which expresses the *analysandum* must be synonymous with any expression which expresses the *analysans*.

Working out an instantiation of this characterization, using the easy and time-honored example, ‘bachelor’, yields the following. The *analysandum*, or “thing to be analyzed”², is ‘bachelor’; giving an analysis involves mentioning ‘unmarried man’ as

¹ ‘Conceptual analysis’, in the sense I use it here and in what follows, refers specifically to the development of analytically justified philosophical theories through the analysis of concepts. Other philosophical thought, which can legitimately, if more loosely, be called ‘conceptual analysis’, attempts to work out consequences of and identify inconstancies in our use of and thinking about various terms of import. This latter sort of enterprise may or may not count as conceptual analysis in my specific sense; whether it does will depend on the nature of the justification offered for such projects.

² ‘-and-’ is the neuter future passive participial suffix stem in Latin. Thus, ‘Amanda’ – she who is to be loved; ‘Miranda’ – she who is to be admired, and so forth.

the *analysans*, the “analyzing thing”. Suppose an entity in question is W. H. Auden³. ‘Unmarried man’ as the *analysans* is such that a) nobody can know that Auden is a bachelor without knowing that Auden is an unmarried man; b) nobody can verify that Auden is a bachelor without verifying that Auden is an unmarried man; and c) any expression that expresses ‘bachelor’ is synonymous with any expression that expresses ‘unmarried man’. Other accounts of analysis have been offered, but this one incorporates the various substances of them in that it incorporates appeals both to epistemic (a) and b)) and linguistic (c)) aspects of analysis. Notice, however, that the linguistic criterion is the strongest of the three. It is the strongest, because, unlike a) and b), c) underwrites bidirectional inferences between instances of bachelorhood and unmarried men. Thus, on a) alone, I might, for example, know that Auden is an unmarried man but be unable to infer that he is therefore a bachelor. On a) alone, I can infer from the fact that Auden is a bachelor that he is an unmarried man, but not the other way around. However, according to c) (assuming ‘unmarried man’ denotes a satisfactory *analysans* for ‘bachelor’), I can infer in both directions: given that Auden is a bachelor, he is an unmarried man, *and* given that he is an unmarried man, he is a bachelor. The third criterion, c), legitimates this bidirectionality, because it implies that ‘Auden is a bachelor’ is synonymous with ‘Auden is an unmarried man’, and synonymous expressions are supposed to be logically equivalent.

‘Bachelor’ marks what looks like a very easy case and nicely exemplifies how

³ Auden is a nice example here, because he was gay and was involved in a spousal long-term relationship with Chester Kallman, which Auden himself referred to as a “marriage”. Intuitions, thus, are likely to differ as to whether he was a bachelor, given that he considered himself to be married but was not legally married.

clear and explicit the products of such analysis would be (if we overlook the fact that cases like Auden advise caution as to just how reliable such an analysis could be even for “easy” examples). Once we extend the scope of analysis from what many see as philosophically uninteresting topics, like bachelorhood, to include traditional philosophical topics, this sort of analysis is thought both to represent ideal philosophical method and to be the especial domain of philosophy. This is precisely because of the use of analytic truths (statements that are true in virtue of synonymy relations between terms) as either evidence or summations. Such truths are also taken (in various senses) to contrast with empirical reasons and empirically justified claims. Thus, analyticity is supposed to be one of philosophy’s answers to the use of observation in science and a sure foundation for the sorts of claims and theories philosophers make. Where scientists use language to report (and generalize) observed facts about the world (synthetic truths), philosophers are supposed to rely on truths that are true irrespective of how the world happens to be; such truths are often described as analytic, or, in the prevalent formulation: true in virtue of meaning alone.

In this chapter, I will argue that philosophy, conceived as a non-empirical discipline drawing on the authority of analytic truths, is *ill* conceived, because the analytic/synthetic distinction should be rejected, at least any version of the distinction significant enough to ground interesting philosophical claims. Thus, even if there is *some* distinction that might reasonably be called an ‘analytic/synthetic distinction’, no examples of statements designated as analytic according to that distinction are interesting enough to underwrite worthwhile philosophical

elaboration. Without analyticity there can be no analysis, in the traditional sense characterized by Moore. I will discuss why the analytic/synthetic distinction should be rejected; in the next chapter, I will develop and recommend an alternative to analysis: explication, which does not rely on the untenable distinction between analytic and synthetic statements and promises to pay philosophical dividends in ways that traditional analysis has failed to do.

II. Against Analyticity

A. Outlining

In the argument that follows, I rely on, elaborate, and variously depart from three main confederates: W. V. Quine, Nelson Goodman, and Hilary Putnam. In proceeding, I will discuss elements of each of my confederates' views, incorporating rejoinders to the counterarguments of others as well as clarifications and counterarguments of my own, where needed. First, I will give an overall impression of how I see the issue.

Quine's attack on analyticity, often wrongly attributed primarily to his holism⁴, actually focuses on issues of explanatory poverty: analyticity is used often by philosophers as an explanation of some other thing – the truth of a certain class of sentences – but the explanation itself raises many more difficult philosophical questions, none of which have been outfitted with satisfactory answers. If the terms in which analyticity is explained are both philosophically robust and lack an explanation, then analyticity hasn't been satisfactorily explained. In this way, the

⁴ In, for example, Müller (2012)

rejection of analyticity appeals broadly for a sound argumentative structure: contentious claims ought to be supported by claims that are uncontroversially acceptable or claims that can be demonstrated to be true by some public and acceptable procedure of verification.

A sort of meta-lesson of the rejection of the analytic/synthetic distinction in Quine's philosophy (as well as Goodman's and Putnam's) follows a lineage back to Frege's distinction between *Sinn* and *Bedeutung*. In his observations that meaning cannot be adequately accounted for in terms of reference alone, Frege incorporates facts about an expression – facts other than that it refers to such-and-such – into the meaning of that expression. Thus, restating, meaning (and, for sentences, truth) involves a “world contribution” and a “word contribution”. By introducing the notion of *Sinn* as the partner of *Bedeutung*, it's clear that Frege was worried about the “word contribution” getting included. Quine's rejection of analyticity aims to correct the reverse tendency: just as the truth of a sentence is never understood in abstraction from the actual language in which that sentence is formulated, neither is it ever understood in genuine isolation from contact with the empirical world. Yet precisely the latter characterizes all principle attempts to define analytic statements. Analytic statements, characterized as statements that are true in virtue of meaning alone, assume a notion of meaning whereby the empirical, or “world contribution” can be excised from bits of language.

While many philosophers reject Quine's arguments on substantial grounds, motivated by an unwillingness to give up the sort of warrant provided by appeals to analyticity, others have also pointed out that Quine's rejection of analyticity, in some

sense, seems to miss its mark, insofar as it aims to undermine Carnap's notion of "analytic-in-L" for non-variable L.⁵ In the context of Carnap's formal language systems, 'analytic-in-L' denotes a class of sentences that are true according to a systematic stipulation of the system, which may be more or less arbitrary and made as a matter of policy when the rules of the language are formulated. Given this, Carnap's 'analytic-in-L' should be seen as an explication of the ordinary term 'analytic' in a systematic context, not an analysis or clarification of the ordinary concept of analyticity; and, more importantly, 'analytic-in-L' is not a term that can do any of the work to which philosophers want to put 'analytic' (without, at least, begging all of the philosophical questions by incorporating the right "meaning postulates" into the rules of L). This situation raises the question of what explication is, what it's good for, and how it's different from conceptual analysis. I will address that topic in Chapter 2.

This indicates an awkward state of affairs. Much effort has been spent "refuting" Quine's arguments on the part of philosophers who accept and want to defend the traditional methodology of analysis. Were there a conclusive refutation of Quine available, however, it would hardly satisfy such interests. This is because Quine's arguments are directed primarily against the explication of analyticity offered by Carnap, and this version (and the systematic philosophy of which it is a part) could do nothing to help underwrite traditional analysis. In some respect, Carnap's view on that sort of philosophical project is even dimmer and his departure from it is more radical than Quine's is. Carnap views the type of disputes

⁵ Examples are in Marian David (1996) and Gary Ebbs (1997).

typical in projects grounded in conceptual analysis as fundamentally *meaningless*, because of the lack of public, shared standards for adjudicating between competing claims. Where Quine would criticize the same disputes as either methodologically unsound or as unscientific, Carnap would just say that they're so much nonsense. To take a familiar example, overcoming Quine's objections to the analytic/synthetic distinction would not help philosophers who want to talk seriously about, say, whether a fetus is a person. Carnap would offer various alternatives for resolving this dispute, such as the following: a language system, L_P , for discussing fetuses and persons where 'All fetuses are persons' is analytic-in- L_P ; and a language system, L_{NP} , where 'All fetuses are not persons' is analytic-in- L_{NP} . The question of whether to adopt L_P or L_{NP} would be for Carnap wholly a matter of pragmatic choice and not something subject to systematic deliberation (which for him is always subject to a set of predetermined and explicit rules for formation and transformation of sentences in a language system). It is obvious that this radically deflationary and anti-philosophical way of resolving the dispute would be wholly unsatisfying to those who take the question seriously.

B. One Dogma, and Another

i. Context

Now that I have criticized others for missing the significance of Quine's arguments against analyticity by taking them out of the context of the debate with Carnap, I want to develop a somewhat decontextualized reading of the same arguments for different purposes. As I suggested above, the main thrust of Quine's

arguments focuses on explanatory force, not on empirical holism. The summary idea is that all efforts to explain analyticity, understood as “truth in virtue of meaning alone”, rely in some respect on concepts that either assume analyticity itself or are explanatorily nebulous. For Quine, however, explanatory clarity should be understood always from a naturalistic point of view; genuine explanations are, for him, always explanations within the framework of empirical science. We might therefore modify my summary characterization of his arguments to say that he criticizes all attempts to explain analyticity because they rely on scientifically unexplanatory concepts. Given this, the arguments he makes will fail to be convincing or even *relevant* to those who do not share Quine’s motivating assumption and dogmatic endorsement of philosophical naturalism. I suggest, however, that we need not be naturalists in order to accept the shape of Quine’s arguments; the notion of explanatory clarity need not be understood specifically in naturalistic terms. Decontextualizing Quine’s arguments in this way, moreover, can help us incorporate the strongest elements of Quine’s, Putnam’s, and Goodman’s rejection of a philosophically useful analytic/synthetic distinction into a coherent picture.

To say that for Quine explanatory clarity is to be understood naturalistically, requires a bit of careful formulation in the particular case of logic, which in turn can help to further clarify the hard distinction between Carnap and Quine’s approaches to philosophy. As suggested before, Carnap takes the resolution of disputes by the formulation of clear and mutually acceptable rules for resolving them to be a task of philosophy; and such rules comprise logic for Carnap. Thus, for Carnap, logic is a

methodological tool for resolving disputes; as such, the formulation of logical laws is not subject to rational (for him, rule-governed) deliberation; it is a matter of purely pragmatic concern. For Quine, however, logic itself is a part of science, and Quine's schematic understanding of logic is explained in terms of theoretically identified logical terms, substitution of non-logical particles, and truth. Valid deductive inferences are just those whose premises, when conjoined as an antecedent, and conclusion, taken as a consequent, form a conditional that is schematized by a valid truth-functional schema. And valid truth-functional schemata are just those that remain true under all interpretations of their non-logical particles. This view of logic fits well with Quine's naturalism, because the clarification of deductive inference that it engenders allows scientists to perform sophisticated deductions that would not otherwise be possible, and it helps clarify the commitments and consequences of theoretical beliefs in every area of science. Furthermore, the machinery of this view of logic, as opposed to other approaches, such as the more prevalent view offered by Barwise and Etchemendy (Barker-Plummer *et al* 2011), which involve theoretical appeals to modal notions, runs more or less under the steam of one powerful concept: truth. Thus, Quine accepts logical truth as naturalistically explanatory because it pervades scientific reasoning and it can be explained in terms of a concept that science itself already assumes.

This provides the background for correctly understanding Quine's arguments in context. In Carnap's view, analyticity should be understood as part of logic, which is just a bunch of rules we settle on to use in resolving disputes. One manner of such rule is a "semantical" rule, which explicitly restricts the range of possible state

descriptions for a language system. Such a rule might, for example, exclude all of the state descriptions in which an object is described as a bachelor and either married or not male. Quine's explicit complaint with semantical rules is that, unlike valid schemata of first-order logic, they are scientifically unexplanatory. Having developed the appropriate context in which to read Quine's arguments as well as suggested with caution at how they might be read in an appropriately decontextualized way, I will now turn to discuss the three of them in turn.

ii. The Definition Definition

As suggested in the previous section, Quine's task in "Two Dogmas of Empiricism" (Quine 1961) is to question whether analytic statements, as a category, can be given an explanatorily satisfying account, especially as compared with logically true statements. The statements he pairs to exemplify these two classes are

(1) No unmarried man is married

and

(2) No bachelor is married.

(1) can be explicated as a logical truth. If we regiment it into a logical language, we get

(3) $\sim \exists x (\sim (x \text{ is married}) (x \text{ is a man}) (x \text{ is married}))$

which is schematized by

(4) $\sim \exists x (\sim Fx \cdot Gx \cdot Fx),$

which is true under every reinterpretation of its only non-logical particles, 'F' and 'G'. The same treatment cannot be provided for (2), which is schematized by

$$(5) \sim \exists x(\sim Fx \cdot Gx \cdot Hx)$$

which is false under any number of interpretations of 'F', 'G', and 'H'. Thus, unlike (1), (2) does not count as logically true under the account based on truth under every interpretation of the non-logical particles.⁶ So, the question is whether, given this, (2) can be explained to be sufficiently like (1) so that the two could count as falling within the common type: analytic statements.

The vague and presystematic characteristic taken to be common between (1) and (2) is that they are "true in virtue of meaning alone" and, hence, devoid of empirical content. The notion that (1) is non-empirical is given a certain degree of precision under the Quinean account of logical truth; the fact that the non-logical parts of the schema can be reinterpreted in any fashion without changing the truth-value of the schema makes it seem as though the empirical contribution to the truth of the statement is nil. This is rather misleading, however, especially from a Quinean perspective, which subsumes logic under the heading of science, most generally, and has it not that logical truths are non-empirical but that they are simply the most general type of empirical truth.

Nevertheless, if we forget for the moment about the cautionary note that construing logical truths as non-empirical somewhat misconstrues Quine and focus on the fact that at least no *particular* empirical facts influence the truth of logical truths, we can see how (1) and (2) are supposed to be alike. One attempt to correlate them has been to urge that *the reason* that no particular facts matter to the

⁶ From Carnap's point of view, however, even this modest claim is bound to appear question-begging, since Carnap understands "Semantical rules" to be part of logic, so he would object to the notion that 'F' and 'G' in (5) are designated as non-logical particles. I will return to this issue below, when I discuss Quine's argument against analyticity defined in terms of semantical rules.

truth of (2) is that it is *made true* by a relation of synonymy between, say, 'bachelor' and 'unmarried man'. 'Bachelor' just means the same thing that 'unmarried man' means, and since no unmarried man is married, (2) no bachelor is married. We start with a logical truth, 'no unmarried man is married', replace an expression, 'unmarried man', with an equivalent one, 'bachelor', and end up with our analytic statement.

Once this move is made, however, it becomes necessary to clarify the claim that 'bachelor' and 'unmarried man' are synonymous. The first appeal towards this Quine entertains is that synonymy is underwritten by definition: we find out which expressions are synonymous by finding out definitions. It is, of course, legitimate in ordinary contexts to offer a definition of a term, 'F', when queried, "what does 'F' mean?" And it's perfectly ordinary, when 'G' is a single term or concise phrase, to paraphrase "F means G" as "'F' and 'G' are synonymous'. This, as Quine maintains, "put[s] the cart before the horse," however (1961: 24). Presumably, a lexicographer formulated the definition, because she determined there to be a preexisting synonymy between the term and its definition, not the other way around. We cannot appeal to definition to explain what synonymy is, when definitions themselves are *based on* synonymies in the first place. This confuses the process of finding out what's synonymous with what with the goal of explaining the relation of synonymy itself. Or, more explicitly, if a term is in fact synonymous with its definition, the term has the definition it does *because they are synonymous*; it's not that they are synonymous because of the definition. Furthermore, the appeal to definition betrays the original commonsense characterization of analytic statements as non-empirical,

because synonymies, understood according to definition are the consequences of empirical investigation. Lexicographers work by empirical observation of language use, not by direct intuition of analytic truths or of synonymies.

This last point segues into Quine's second argument about definition, where he considers the specialized variety of definition he associates with Carnapian explication, because explication represents a departure from descriptive definition, where the objective is simply to provide a summary gloss on the antecedent uses of a term. In this way, also, for those more familiar with the terminology of linguistics than that of philosophical semantics, explication and definition (in the first sense Quine discusses) represent counterparts to prescriptive and descriptive definition, respectively. Prescriptive definition, like explication, can *produce* definitions (and recommend them for adoption) which can (sometimes radically) depart from antecedent usage in certain contexts. Under Quine's characterization, explication involves not merely the reporting and summarizing of pre-existing synonymies but the modification of pre-theoretically vague terms that function well in some contexts. Explication, then, attempts to clear away the vagueness (or other theoretically undesirable qualities) and replace the ordinary term with a systematic, artificially designed one. Furthermore, given the difference in meaning (because of the modification) between the pre-explication term and the new term, there are sure to be contexts in which the new term functions differently from the old. As long as these differences do not interfere with research and inquiry and they are relegated to contexts other than those in which the pre-systematic term functioned well, they fall under the Quinean heading of "don't cares". Although Quine endorses

explication as a philosophical method and even identifies an example instance of it a “philosophical paradigm” (1960: 257–262), he deploys the same argument against explication as an account of synonymy that he deploys against ordinary descriptive definition. That is, although the ways in which the explication differs from the term it replaces do not represent preexisting synonymies, these differences are immaterial (“don’t cares”) and they are permitted only because they affect contexts of use that are unrelated to the theoretical purposes of the explication in the first place. The parts of the explication that *do* function in theoretical contexts do so because the replaced term was itself sufficient *in these contexts*. In this way, the aspects of an explication that matter (in terms of meaning) are based on prior synonymy relations (just like descriptive definition); the parts that aren’t based on such relations (and so do, in this sense, artificially prescribe synonymy) are rather beside the point, because the contexts where they’re relevant are never contexts of use in theory.

Still, if we take seriously the notion of an explication as a *replacement* of an old term with a new one, adopted as a matter of policy, it is not clear why this should not count as the establishment of a genuinely new synonymy relation between the new term, the *explicans*, and its definition, the *explicatum*. There may be considerations of the *explicandum*, the presystematic term being replaced, that influence the choice of *explicatum*, but that does not mean that the relation of synonymy between the *explicans* and the *explicatum* existed presystematically: it is, after all, a *different* synonymy relation from the presystematic one. This sort of critique of Quine’s analysis of the relation between explication and synonymy is

made more serious, furthermore, if we also question his view of “don’t cares”. For some examples of explication, different from Quine’s flagship example of the ordered pair, the new aspects of the explication are precisely what is interesting about them and what pays philosophical, creative, scientific, and technological dividends. Given cases of this sort, the argument that explication relies on old synonymies instead of underwriting new ones becomes much less satisfying, since, for such cases, it is precisely the new contexts that make the explication worthwhile. This sort of criticism, although Putnam does not introduce it himself, fits well with his rejection of a Carnapian analytic/synthetic distinction (Putnam 1975), since it focuses on the introduction, development and use of explications in actual practice. I will say a bit more about this line of critique when I discuss Putnam’s arguments below; but a complete reckoning of this worry will not be fully clear until I discuss the shortfalls of Quine’s account of explication in the next chapter. However, it should be noted at this point that, although I reject Quine’s argument regarding synonymy based on explication, because I think he mischaracterizes the role played by explications in actual practice, I also recognize that this rejection does not help the would-be conceptual analyst. This isn’t the right sort of synonymy to underwrite claims to analytic truth, precisely because such synonymies are newly minted with the constructive formulation of explications and don’t identify *prior* synonymy relations.

In this respect, the new synonymies introduced by explications are similar to the sort of synonymies that Quine admits can be explained by definition: conventional abbreviation of symbolically cumbersome notations by concise ones.

Such abbreviations effectively mandate that, from now on, some concise expression will be synonymous with some complex one. This sort of thing does not come under the same attack as explications do, because they introduce novel notation. Since such notation is novel, its correlation with more complex pieces of notation cannot be charged with relying on prior synonymies, since these new notational items have no prior meaning at all.

The argument now comes to a point where a specific sort of defense seems necessary, but Quine passes it without comment. Quine has been discussing the possibility that analyticity, understood as truth in virtue of meaning, could be explained in terms of synonymy relations. This introduces the need to explain synonymy, one attempt at which is to explain it in terms of definition. He finds two types of definition to *rely on* synonymy rather than to generate it, yet when 'definition' is understood as conventional notational abbreviation, he accepts it as genuinely founding synonymy relations. Placing such cases, then, back within the context of the discussion, the question becomes obvious of whether definition, in the sense of notational abbreviations, could be offered as a justification for a claim that a certain sentence is analytic. Whatever problems we may have in explaining synonymy, meant to underwrite analyticity in more common examples, if definition in this narrow sense does sufficiently explain claims of synonymy, as Quine seems to admit, then it seems that Quine commits himself to the notion that statements identifying such abbreviations with their more longwinded correlates could be justified as analytic in a way that doesn't rely on unexplained notions.

This, like the sort of synonymy I suggested could be generated by explication, will do little to assist the conceptual analyst because the sorts of analytic claims it could be used to justify would be so philosophically banal as to be not worth mentioning. However, within the context of Quine's argument, this opening represents a real problem. Quine doesn't aim to argue (as Putnam does) merely that there are no philosophically interesting analytic statements; he wants to argue that there are no analytic statements at all. Yet, by the criteria of his own argument up to this point, it appears as though there is a space for a class of such statements; in order to show that there are no analytic statements whatsoever, he would have to answer why synonymy in the case of notational abbreviation wouldn't count as a sufficient explanation, given that "definition" in this case fully explains the fact that the expression and its abbreviation are synonymous. This point may seem to just be a picky gripe, but it isn't, because it reveals ways in which Quine's position on analyticity is connected to his naturalism, and, specifically, his behavioristic skepticism about meaning. About languages that include such notations and definitions he says (1961: 26–17):

The inclusive language, though redundant in grammar and vocabulary, is economical in message lengths, while the part, called *primitive notation*, is economical in grammar and vocabulary. Whole and part are correlated by *rules of translation*⁷ whereby each idiom not in primitive notation is equated to some complex built up of primitive notation. These rules of translation are the so-called *definitions* which appear in formalized systems. They are best viewed not as adjuncts to one language but as correlations between two languages, the one a part of the other.

⁷ The emphasis is mine.

The notion that the language containing abbreviations and the “primitive notation” are different languages and that definitions between them are *translations* is apposite. From Quine’s naturalistic perspective, the meaning of sentences is analyzed in terms of speech dispositions under differing conditions of sensory input and “manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions yet incompatible with each other” (1960: 27). Taken in this way, as translations, it becomes clear what Quine would say about the possibility of abbreviation-definitions justifying claims of analyticity; he would say that, such claims are scientifically meaningless, since scientific inquiry has to do with empirical input, and translation is indeterminate with respect to such input. This last line of argumentation is important precisely because it indicates the extent to which Quine’s naturalism pervades his view of these issues.

It also indicates the degree to which criticisms like that of Scott Soames, which assert that Quine is inconsistent on synonymy, are misplaced. Discussing an elaboration of the arguments about analyticity Quine published fifteen years after “Two Dogmas of Empiricism” in *Philosophy of Logic* (Quine 1986), Soames claims that

It is amusing that Quine indicates that *cordate* is short for *creature with a heart*. What he means, of course, —though he doesn’t say it—is that as he uses these terms, they are **synonymous**. But if that is what he means, then there must be such a thing as synonymy after all. So his very example seems to presuppose the position he uses the example to argue against (Soames 2003: 364).

As I have urged throughout my discussion of Quine’s first argument, his concern with analyticity as a philosophical concept is that it is explanatorily impoverished; it

relies on notions, such as meaning, necessity, and *synonymy* that are themselves in need of explanation and so cannot be legitimately used in a philosophical theory without themselves being explained in some non-question-begging way. This is perfectly consistent with some varieties of synonymy both existing and being detectable. One such variety is involved in the introduction of novel notation; that type of synonymy is fully explained, because it exists wholly as a consequence of the conventional definition. In this sense it is purely a matter of convenience, and any philosophical claim grounded on it is bound to be trivial and uninteresting. Also, if we look at 'cordate' as what it is, the introduction of novel notation as a translation of the syntactically more complex expression 'creature with a heart', we see that the amusement Soames reports at having identified an inconsistency in Quine's view is itself cause for a bit of a chuckle.

iii. Interchangeability: Chickens and Eggs, in the Absence of Biology

I ended the previous section by elucidating a confusion of Scott Soames' reading of Quine. This served to reiterate my point that the overall structure of Quine's argument focuses on explanatory sufficiency. It also serves to introduce Soames, whose "refutation" of Quine's circularity argument against analyticity has become influential; furthermore, Soames' supposed way around these arguments has been seen as reintroducing the possibility of traditional philosophical analysis. Although Soames presents himself as defeating Quine's *arguments* against the analytic/synthetic distinction; this could not be correct even if nothing else Soames says is untrue, because he only discusses the second argument, dealing with

interchangeability. This fact is stranger still, given that he advertises his discussion (and admonishes his reader) that it should take the context of “Two Dogmas of Empiricism” into account. In this, at least, he is correct. The context he means to include is that of a response to logical positivism. And this is also correct, but there is positivism and there is positivism. Perhaps the argument regarding interchangeability addresses some of the less sophisticated positivists, like Ayer, but the argument about interchangeability is wholly immaterial to Carnap’s notion of analyticity, which Quine explicitly addresses. The kinds of questions left open at the end of Quine’s argument about interchangeability, like how to define necessity without assuming analyticity, are answered directly (as well as conventionally and trivially) in a Carnapian language system by semantical rules. The fact that Soames takes himself both to have refuted Quine and provided the appropriate context for understanding his arguments coupled with the fact that he nowhere so much as mentions the notion of semantical rules is altogether puzzling. Despite the problems with his overall picture these remarks indicate, Soames’ discussion of the interchangeability argument is worth considering, for its insights and errors, and because both have been influential.

So, to pick up the thread of Quine’s argument: trying to develop an explanation of what it might mean to say that a sentence is true in virtue of meaning, he has considered that it could mean that such a sentence can be transformed into a logical truth by replacing synonyms for synonyms. This requires explaining synonymy; one idea is that synonymy is explained in terms of definition, but it has turned out that, in all but one trivial sort of instance, synonymy explains definition,

not the other way around. Quine identifies another tack for explaining synonymy as coming from Leibniz: interchangeability *salva veritate*⁸.

Defining synonymy as interchangeability *salva veritate* looks promising for obvious cases:

1) Willard is a bachelor.

is true just in case

2) Willard is an unmarried man.

is also true. Since substituting 'unmarried man' for 'bachelor' in 1) preserves the truth-value of 1), this could be grounds for asserting that 'bachelor' and 'unmarried man' are synonymous. But the problem with this approach is that, while it reaffirms many of our presystematic intuitions about what's synonymous with what⁹, it contrasts sharply with other such intuitions in ways that seem to be unacceptable. For example:

3) The number of books in the *Odyssey* is even.

is true *just in case*

4) The number of hours in a day is even.

is true. 4) is generated by substituting 'the number of hours in a day' for 'the number of books in the *Odyssey*' in 3). Since both expressions denote the number twenty-four, they both could be paraphrased as 'Twenty-four is even', and, of course, twenty-four is even *just in case* twenty-four is even. Thus, we have a case of

⁸ *Salva veritate*: with truth preserved

⁹ I should say, though, that I do not share the intuition that 'bachelor' and 'unmarried man' are synonymous, and I will argue this point below, when I discuss Putnam's arguments against a philosophically interesting analytic/synthetic distinction. I use this example only because Quine, Putnam, Soames, and everyone else I know of seem to accept it.

interchangeability *salva veritate*, but the prior notion that ‘the number of hours in a day’ is synonymous with ‘the number of books in the *Odyssey*’ is unintuitive and lacks motivation not related to the view. There are plenty of people who are old enough to tell time that have never heard of the *Odyssey*; though they may grasp the expression ‘number of hours in the day’ they do not thereby grasp the expression ‘number of books in the *Odyssey*’. Just as it required an act of discovery, or realization of a coincidence, that the morning star and the evening star are one and the same, so does it require discovery to know that the number of books of the *Odyssey* is equal to the number of hours in a day.

It might be objected to this characterization that it confuses the meaning of an expression with knowing the conditions under which it can be truly asserted. This is apposite in some contexts, but the current context is not one. Since the discussion of synonymy is as a possible way of explaining analyticity, and analyticity is characterized as truth in virtue of meaning and truth irrespective of the facts, the conditions of true assertion are integral to the discussion. Thus, even if it were argued that ‘the number of books in the *Odyssey*’ means the same as ‘the number of hours in a day’ by claiming that knowledge of assertion conditions is not necessary for knowledge of meaning, this would just entail that synonymy is a bad explanation of analyticity. Surely, if we think analytic statements are those statements that are true irrespective of the facts, then ‘The number of books of the *Odyssey* is equal to the number of hours in a day’ would hardly count as analytic, even if we stubbornly held that ‘the number of books of the *Odyssey*’ and ‘the number of hours in a day’ are synonymous as a consequence of the interchangeability criterion.

Given this problem, that some expressions can be interchanged for one another without meaning the same thing in any ordinary sense, the definition of synonymy as interchangeability *salva veritate* stands in need of modification. Quine suggests that this issue would be fixed by including, in the set of sentences to be tested for interchangeability *salva veritate*, those of the following type:

(5) Necessarily, all and only bachelors are bachelors.

Inclusion of intensional language, like ‘necessarily’, is supposed to assist with the definition of cognitive synonymy based on interchangeability *salva veritate*, because a statement like

(6) Necessarily, all and only bachelors are unmarried men.

looks *prima facie* true, or at least it has a better chance of being true than, say

(7) Necessarily, the number of books of the *Odyssey* is equal to the number of hours in a day.

It *might* appear unlikely that a scenario can be coherently described involving a bachelor who is either not a man or not married, but we can easily imagine scenarios in which the *Odyssey* didn’t have twenty-four books or in which days didn’t have twenty-four hours. It might turn out, for example, that the *Telemachy* is apocryphal, so the *Odyssey* only has twenty books.

However, to say that this move is to *add* intensional contexts to those in which the *salva veritate* question is to be asked is importantly misleading. If two terms are interchangeable *salva veritate* in contexts including ‘Necessarily, ...’ then they will be interchangeable in ordinary extensional contexts as well. And this indicates the problematic character of the move: if the terms are interchangeable *salva veritate* in contexts of necessity, this serves to explain the fact that they are

interchangeable *salva veritate* in purely extensional contexts as well. Thus, the explanatory work is being done more by necessity than by the notion of interchangeability *salva veritate*, so in order to understand the account of synonymy based on interchangeability, it is necessary also to understand necessity. One way in which we might approach the latter question is in terms of analyticity: necessary truths, such as (5) and (6), are analytic. That this won't help in the current project is beyond doubt: the whole reason we wanted to understand necessity in the first place was as an aid to understanding analyticity, and we therefore cannot appeal to analyticity in our account of necessity.

Quine sums up his rejection of the argument from interchangeability with a telling quip: "Our argument is not flatly circular, but something like it. It has the form, figuratively speaking, of a closed curve in space" (1961: 30). In the logical sense, of course, the argument, as characterized, *is* circular. 'Flatly', modifying 'is-such-and-such', usually indicates that the predicated quality is clearly exemplified, that there is no question about the object's being a such-and-such. But the first clause is also ambiguous as to what 'flatly' modifies; whether it modifies 'circular' or 'is circular'. The latter sense is more natural to the context, since what is being talked about is an argument, but the former sense is relied on in the second sentence. That is, a flatly circular thing – a thing with the shape of a flat circle – is contrasted with a "closed curve in space". Given that space is *not* flat, opposing the two makes sense. As a "closed curve" is one way of describing a circle, but if that circle is in space, it will not be "flatly circular". The pun may indicate an important point in support of Quine's view. First, with 'flatly', 'circular', and 'closed curve', it

exemplifies the fact that words are used in ways that vary from context to context and change with theoretical development. It used to be the case that one could infer from the fact that a thing is flatly circular that it is a closed curve in space, and vice versa; this is not the case any more. The difference is that we no longer believe, as we once did, that space is flat, so closed curves in space need no longer be flatly circular. The illustration of this change is apposite for the issue at hand, because the modeling of physical space on the axioms of Euclidian geometry underwrites the illicit inference from 'flatly circular' to 'closed curve in space', and precisely this modeling has previously been believed to be analytically true. The pun, therefore, serves to cast doubt on the notion that analyticity could be scientifically useful.

Soames, who pretends that the circularity argument is Quine's only argument against analyticity, offers a vigorous and triumphalist counterargument. His basic position is that the circularity argument works only if two theses of positivism are assumed (2003: 360):

- T1: All necessary (and all apriori) truths are analytic. (For all sentences S, if S expresses a necessary (apriori) truth, then S is analytic.)
- T2: Analyticity is needed to legitimate and explain necessity (and aprioricity).

Soames' attention to the fact that we can correctly read Quine's arguments only with an eye to their historical context issues wholly in the claim that Quine held T1 and T2 along with the positivists he was arguing against. This mischaracterizes the historical context of Quine's argument in important ways that I've already indicated. Furthermore, textually, it seems clear that Quine accepts T1 only hypothetically for the sake of this particular argument; he does not accept T2 at all, however; and

supposing he does accept it obstructs a correct reading of his argument, which is an argument about explanatory sufficiency. Outside of the context of a specific argument, Quine is likely to deny that he understands either T1 or T2, so it's not developing the historical framework to say that he accepts T1 and T2, it's mischaracterizing Quine, his relation to the positivists and his naturalistic conception of philosophy. That Quine assumes T1 only for the sake of the circularity argument is evidenced when he says 'Necessarily, all and only bachelors are bachelors' "is evidently true, even supposing 'necessarily' so narrowly construed as to be truly applicable only to analytic statements" (Quine 1961: 29). This indicates that Quine is himself not committed to any particular view on necessity and its relation to analyticity.

However, having corrected Soames in his claim that Quine holds T1, let us look at the role that T1 plays in Quine's argument, getting clear on this role will help us address T2 as well. Quine assumes T1 for the sake of the argument as well as its converse: all analytic truths are necessary. This is because the identification of analyticity with necessity was a prevailing view, not because it is correct or taken by Quine to be correct. From a pretheoretical perspective, this equation is intuitively compelling, analytic truths are supposed to be true in virtue of meaning alone, and thus true regardless of what the facts happen to be; necessary truths are supposed to be true without possibility of being false, thus: true *regardless of what the facts happen to be*. If both types, analytic truths and necessary truths are true regardless of what the facts happen to be, it is difficult to see how a sentence could be one without being the other. However, this sort of consideration would be, for Quine, the

starting point of philosophical inquiry, not anything like a conclusion. He is trying to make sense of the notion of analyticity; necessity comes into consideration as a tool for repairing the explanation of synonymy in terms of interchangeability *salva veritate*, but it turns out that we don't have a grip on necessity that doesn't assume analyticity. Soames' complaint is that we *do* have such a grip, that Quine is wrong to assume all necessary sentences are analytic or that we need analyticity to explain necessity. This is a consequence of Soames' reading of the twentieth century as one in which Quine overcomes positivism, and Kripke overcomes Quine, giving a new legitimacy to the notion of necessity, understood without reference to analyticity or meaning at all.

That there are other ways of viewing the contribution of Kripke and the trends of late twentieth century and contemporary philosophy should be beyond doubt. However, such a reconstruction is immaterial to a discussion of Quine. If we had an independent grip on necessity, as Soames thinks we do, then Quine would accept it as explaining synonymy in terms of interchangeability *salva veritate*. But, the problem with all such explanations is that they are unlikely to be scientifically meaningful for Quine. Given this, we should reject T2 even as an argumentative assumption of Quine's. It's not that he thinks analyticity is needed to explain necessity. It's that *something* is needed to explain necessity; if we *had* a scientifically meaningful understanding of analyticity, then that would do, but that doesn't entail that, according to Quine, analyticity is the only way to explain necessity. So just as the arguments about definition had, fundamentally, to do with explanatory sufficiency, where a thing is sufficiently explained just in case it is explained in terms

we already understand, so too is the circularity argument. And about the Soames-move to explain necessity in terms of something else, it's doubtless that Quine would reject such an explanation as not scientifically meaningful. Thus, finally, with the circularity argument as well as the others, the most basic motivating assumption is philosophical naturalism, and the latter represents the measure of explanatory sufficiency for Quine.

There remains one argumentative thread left loose from the skirmish between Soames and Quine. Above, I claimed that, were some (naturalistically) viable explanation of necessity available, Quine would count it as legitimately grounding an account of analyticity. Soames, however, thinks that even this is going too far; and it's precisely for this reason that he wants to assert that Quine actually holds T1 and T2, instead of accepting (as the text clearly suggests) that Quine considers these hypothetical assumptions at best. Thus, Soames considers an alternative account of synonymy that highlight what are, for him, the fundamental differences between analyticity and necessity; on Soames' view, Quine would be wrong to define analyticity in terms of necessity, even if he had an independent account of necessity he could accept. Examples Soames gives in this respect are telling (2003: 362):

(8) It is a necessary truth that all and only *equilateral triangles* are *equiangular triangles*.

(9) It is a necessary truth that $2^{10} = 1024$.

Soames appeals to these examples as obvious cases of necessity without synonymy (between the italicized terms, respectively). A Quinean (or even just a philosopher unimpressed with what Soames characterizes as the overwhelming contemporary

consensus) would face little difficulty in dispatching examples of this sort as showing that we have a grip on necessity independent of synonymy and analyticity. (8) is not serious because it's only true under some conceptions of geometry; and the choice of a geometry to model space is a matter of empirical inquiry. It's not at all difficult to imagine funny, curved triangles whose angles are equal but whose sides are not. Thus, as an example of necessity, it's poorly chosen, because it is necessary only with respect to a certain set of assumptions, the latter of which are not necessary. As for (9), Quine would probably call the exponent a weathered piece of "novel notation" that creates economy and flexibility and thus counts as grounding synonymy in the artificial sense discussed above in section B.ii.

Considerations of this sort also serve to highlight ways in which we might appropriate Quine's arguments for rejecting analyticity (and possibly necessity as well) without assuming his dogmatic naturalism. As a precissification of the supposedly non-empirical character of analytic statements, appeal to necessity would be useful. But the latter remains unexplained by my lights, even though I do not endorse Quine's naturalism, and despite the examples such as those identified by Soames that are supposed to count as "evidence" of necessary truths that are not analytic. We need not, therefore, construe 'explanatory' as 'scientifically explanatory' in order to commiserate with Quine's confusion over analyticity and necessity.

iv. Ruling out Semantical Rules

Although section 4 of "Two Dogmas of Empiricism" generates discussion less

often than does the holism of sections 5 and 6 or the circularity of section 3, the argument of section 4 on semantical rules addresses the analyticity of logical positivism at its most coherent, in the philosophy of Rudolf Carnap.

Having seen that synonymy, explained variously in terms of definition and interchangeability, failed to elucidate analyticity in an explanatorily satisfying way, Quine directs himself anew to the question of analyticity from the starting point that the meaning of ordinary terms (as had been the focus up to this point) may not be precise enough to ground an account of analyticity. Thus, he considers whether

(1) Everything green is extended.

is analytic. Quine claims to not know whether (1) is analytic, despite the fact that he understands the meanings of 'green' and 'extended'. The idea is that, if we had a more precise language, then it would be clear what's analytic and what's not, both because predicates like 'green' and 'extended' would have precise assertion conditions and because 'analytic', were it a term in a precise language, would as well. It is the formulation of exactly this type of precise language, dedicated to the resolution of "fruitless disputes" that Carnap understood to be one of the principle tasks of philosophy. Thus, the suggestion of Quine's, that it's not analyticity but the language we're using to discuss it, should be seen as drawing focus on Carnap's particular understanding of analyticity.

Before launching into a discussion of Quine's argument, though, since I've given a brief gloss on Carnap's approach to philosophy and the role he thinks precisely specified language systems do and ought to play, it will do to comment in

this context about Soames' attribution of his theses T1 and T2¹⁰ to the positivists. Insofar as Carnap represents the positivists, the positivists don't accept T1 and T2 in any non-relative sense. Carnap may see fit to accept the two theses, insofar as the semantical rules of a formal system imply them or the they turn out to be empirically true under some set of coordination rules that map ordinary speech onto a formal language. But holding T1 and T2 true in either of these conventional senses hardly amounts to the same thing as asserting them as philosophical theses, which is how Soames understands them. To clarify this error at this point might be picky, were it not for the fact that it serves to develop the context of Quine's argument against Carnap by indicating how *non-philosophical* his commitment to analyticity is. This fact will come back into play below, when I argue that Quine's arguments fail to undermine Carnap's position unless Quine's naturalism is assumed.

Up to this point, I have offered some cursory remarks about Carnap's notion of analyticity. In order to understand Quine's criticism of it, as being underwritten by semantical rules, I'll now sketch the view. In *Meaning and Necessity*, Carnap explicates analytic-in- L_2 as a species of logical truth or L-truth in L_2 . L-truth is, in turn, explicated in terms of state-descriptions in a given language system (Carnap 1956: 222ff). Consider, for example, a language, L_1 , containing three predicates, 'F', 'G', and 'H'; two individuals; the quantifiers and the logical operators. A state-description in L_1 is a conjunction of atomic sentences where, for each predicate, each individual is either in its extension or not in its extension, such as

¹⁰ P. 24, above.

$$(2) Fa \bullet Fb \bullet Ga \bullet \sim Gb \bullet \sim Ha \bullet Hb$$

Many statements of L_1 are consistent with (2), such as

$$(3) \forall x(Fx \vee \sim Fx)$$

$$(4) \forall x(Fx \rightarrow (Gx \vee \sim Gx))$$

$$(5) \forall x(Hx \leftrightarrow (Fx \bullet \sim Gx))$$

Regardless of what complete state-descriptions we formulate, (3) and (4) will remain true, but, unless there are restrictions on how the state-description can be changed, some state descriptions will be inconsistent with (5). Thus, for Carnap, (3) and (4) would be L-true in L_1 ; they are true in every state-description of L_1 . Carnap extends the notion of L-truth to cover analytic sentences as well by the introduction of meaning postulates. (5) is a candidate for such a treatment; suppose, for example, that 'H' denotes bachelors, 'F' denotes men, and 'G' denotes married (people). Were the codifier developing L_1 , upon reflection about the meaning of 'bachelor' to determine pragmatically that all bachelors are unmarried men, she could add (5) as a meaning postulate of another language L_2 . This second language can be thus defined as having a narrower range of state-descriptions in the following way: The set of state-descriptions of L_2 contains the same members as the set of state-descriptions of L_1 in which (5) is true (Carnap 1956: 8–10). Restricted in this way, the L-truths of L_2 (still understood as the sentences of L_2 that are true under every state-description in L_2), are the set of logical truths plus the explicitly codified analytic truths; in this case, just (5): interpreted, 'All and only bachelors are unmarried men'.

Quine's argument against this way of explicating analyticity is that it doesn't

explain anything about *analytic* statements, how they could be said to be “true in virtue of meaning alone”. This is illustrated by the fact that nothing in the account excludes the possibility of choosing “meaning postulates” having nothing to do with meaning. I might pick any true sentence, or set of them as restricting the state descriptions in a language system. For example, suppose there is a class of truths that I’m really, really certain about, so I know that restricting my state-descriptions to exclude their negations would never hamper the progress of inquiry; instead of extending L-truth to include truths that are analytic-in- L_2 , I might extend it to include truths that are *super-sure*-in- L_2 . Reinterpreting ‘H’ as ‘is a human being’, ‘F’ as ‘less than 500 feet tall’ and ‘H’ as ‘has counted each number up to $\aleph_0 + 1$ ’, then (5) would schematize a *super sure truth*, but not an analytic one in the sense philosophers usually use that term.

Quine’s worry with semantical rules (or, in the terminology of *Meaning and Necessity*, meaning postulates) is that they fail to pick out anything especially *semantical*, which was supposed to be the characteristic underwriting analytic truth. The super-sure-truths are not true because of semantics: it’s not part of the *meaning* of ‘human being’ that no human beings are over 500 feet tall or that none have counted to $\aleph_0 + 1$. It cannot be consistently replied that the super-sure-truths cannot be included as a constraint on the state-descriptions of L_2 , because such constraints can only be made by semantical rules. This would necessitate the explanation of what semantical rules are, of course, and within Carnap’s system, they are just a set of rules that constrain state-descriptions. Semantical rules are included, as Carnap says, as a matter of the choice of the language system-builder

and as a consequence of her knowledge or belief about what's incompatible with what. But this choice is made outside of the context of a language system, and so it is not meant to reflect non-system-relative truths about synonymy, analyticity, or even semantics, broadly construed. Semantic truths for Carnap, such as they exist at all, are specified and determined conventionally according to the rules of a language system.

The problem with Quine's argument against Carnap is that Carnap accepts Quine's observations about language systems and even the conventionality of 'semantical' in 'semantical rules'. For Carnap, however, what Quine reads as a mark of the unexplanatory character of semantical rules, is simply a feature of the fact that rational disagreement between interlocutors is possible only when a set of rules is available for resolving disputes. For Carnap, we could resolve the dispute with Quine, *if* we had some rule in place to which we could appeal to determine whether a rule is *genuinely semantical* or not. Suppose we were faced with the question of whether the *super-sure*-truth, (S) 'No human beings are over 500 feet tall or have counted to $\aleph_0 + 1$ ' is semantical. Quine would surely regard this truth as not-especially-semantical; thus, his argument against "semantical rules" is that no element of the theory of language systems exclude such examples as analytic truths based on semantical rules, so "semantical rules" don't narrow the field of truths to any set of truths having to do with semantics. However, Carnap would disagree with this assessment on the grounds that a presystematic judgment that S is not true by semantics lacks a rational foundation, because it is not made in accordance with the rules of a language system wherein the difference between the semantical and the

non-semantical is well-defined. Thus, Carnap might propose, in a measure to humor Quine's antecedent philosophical intuitions about the semantic and non-semantic, that the dispute be resolved in a meta-language wherein rules are in place for determining a generalization's status as semantical or non-semantical. If the rules for use of 'semantical' in this meta-language determine that S is not semantical, not a "semantical rule" and not an analytic consequence of such a rule, then S would not be "semantical" in the object language. This is to say, then, that when Quine objects, in the manner of the counterexample, to Carnap's definition of analyticity by offering classes of truths (the "K truths") which *could* be analytic in some Carnapian language systems but which we have no *prior* reason to suppose are analytic, Quine completely ignores the fundamental commitment of Carnap's philosophy. That is, rational determinations about philosophical matters, such as what is and what isn't analytic, are to be made only within the framework of a formally defined language system. Quine's antecedent suspicions that the "K truths" or that the "super-sure truths" are not analytic and don't have anything to do with semantics might be *used* when formulating the language system so that we formally exclude such truths from the analytic camp when we lay out the details of the language system, but such decisions are not made as a consequence of rational determination; they are made as a matter of policy whereby we choose (or not) to use a language system more or less in keeping with antecedent judgments about philosophical matters, such as the question of analyticity.

So, it seems that Quine's criticisms of Carnap more or less fail to meet Carnap on his own grounds. While Carnap takes one of the tasks of philosophy to be the

formulation of languages of inquiry whereby investigators can be sure they are using the same assertion conditions and rules for determination and inference, Quine's criticisms of Carnap's view are always outside a specific language system, where, by Carnap's lights, such criticisms lack salience because they are not subject to public and agreed upon rules of logical significance. Quine's refusal to play according to Carnap's rules, however, is obviously not due to his ignorance of them. Rather, where Carnap's motivating principle is to find ways of resolving the disputes of investigators and interlocutors by placing them on a common, publicly articulable ground, Quine's fundamental philosophical commitment is to an austere naturalism: the only legitimate explanations are explanations within the domain of science. From this naturalistic perspective, Carnap's language systems are meaningless, because they are not part of natural science, which, for Quine, is radically empirical. Thus, while Carnap is prepared to move the boundary of the "semantic" as freely as the adopted language system requires, Quine chooses to settle the question of semantics in a radically empirical way. For him, meaning is analyzed in terms of speech dispositions under various stimuli, and various possible ways of translating a speaker's language to preserve all of the speech dispositions, so he concludes that as far as science goes – as far as generalization over the empirically observable goes – linguistic meaning itself, with all of its niceties, is scientifically and, therefore, philosophically meaningless.

Thus we meet a pretty pass. Clarified in the manner of the foregoing, neither Carnap nor Quine's position looks at all attractive. To appropriate a Quinean appropriation, there are more things in heaven and earth than are dreamt of in

naturalism's philosophy. Yet, the artificial resolution of disputes by formulation of languages to clarify their claims is bound to deeply fracture discourse, to favor one side of the dispute, or simply to remain forever in skirmishes at the formulation stage. There is something deeply unsettling about "Conservapedia", a current answer to the "liberal bias" of Wikipedia, where entries are advertised as (unlike Wikipedia's) hostile neither to the United States nor its religion, Christianity. Yet, with the modification of self-consistency, formalism, and rigor, something very much like Conservapedia could be developed as a Carnapian language system. So, while I do not side with Quine in that the deepest points of his rejection assume his naturalism, if we take the form of his argument regarding semantical rules and apply it to Carnap's language systems-approach broadly, we get a strong criticism. Where Quine complained about the explanatory significance of "semantical rules" because they can cover rules that don't look at all semantic, we might complain alternatively about some language systems that could be formulated in a Carnapian guise; some such systems for resolving disputes and engaging in rational inquiry could themselves incorporate so many vastly erroneous notions that the very idea that they are mechanisms for rational inquiry is preposterous. Suppose we developed such a system by drawing on (and making formally consistent) the ideas of a core of entries in Conservapedia, say, the entries on the United States, church and state, the US Constitution, natural law, marriage, sex, Islam, Darwin, James Baldwin, and liberal. The fact that the formulation of such a language system would be made only as a matter of policy and at the behest of those who want to use it does not undermine the antecedent fact that such a system could hardly be used for

anything calling itself rational inquiry. But, beyond that, the Conservapedia example is a case in point, precisely because it *does* function loosely like a Carnapian language system; it is an alternative compendium of knowledge its proponents develop and advocate on account of what they perceive to be illicit assumptions of the alternative by agreeing to ground rules for interlocutionary discourse as a matter of policy. They don't like the "rules" of the more common web encyclopedia, such as that it "denies" that Jesus' birth is the origin of the current year by using C.E. and B.C.E. instead of B.C. and A.D. Moves of this sort represent fragmentation and the avoidance of discourse more than they do resolution of disputes; when we can "resolve disputes" simply by defining things in the way we like, we're likely to avoid lots of bad and fruitless philosophy (which was one of Carnap's aims), but we're also likely to eschew the difficult work of interlocution altogether, preferring to speak incompatible languages and settle smugly into our dogmatisms, and there are indeed worse dogmatisms than the two Quine decried or the one to which he was devoted.

C. Goodman, 'Grue' and Analyticity.

In the previous section, I reviewed in detail Quine's arguments against analyticity in "Two Dogmas of Empiricism" and placed them within the historical context of the debate with Carnap. In the discussion, it was shown that even if Quine's arguments against Carnap were not accepted, the notion of analyticity that Carnap advocates is the wrong sort of notion to underwrite substantial philosophical claims of necessity, apriority, or analyticity. Whatever truths end up

being analytic in a Carnapian system will be analytic in only an extremely deflated way and will reflect more about the way the speaker has chosen to talk and think than about any substantive philosophical truth.

In my discussion of Quine's arguments, I concluded by maintaining that his position on analyticity is ultimately underwritten by his dogmatic naturalism, which I find unattractive. Nevertheless, if we look at his arguments abstractly, they still have merit and point in a promising direction, that is, if we look at them as arguments questioning the explanatory satisfactoriness of claims to and about analyticity, they are, like all points in Ohio, useful starting points. A philosopher who shares all of Quine's skepticism about the explanatory capacity of appeals to analyticity, meaning, and synonymy but none of his dogmatic naturalism is Nelson Goodman. Although Goodman's alternative to naturalism, pluralism, culminates under his tutelage in some weird and contrary-to-ordinary ways to talking, such as about "worldmaking", Goodman's approach does manage in the particulars to maintain a respect for ordinary, mundane truth ascriptions that don't make the scientific grade without being metaphysically or semantically inflationary *a la* analyticity, necessity, apriority, and their storied fellow travelers.

Although Goodman's work is as much respected as it is maligned, the deep affinity his work shares with Quine's skepticism about analyticity has remained largely unnoticed. This is especially striking, given that the point in Goodman's philosophy where the relation is perhaps most readily perceivable is in his "new riddle of induction", which is also his most broadly acclaimed and widely discussed offering. Although the motivation and defense of Goodman's 'grue' paradox share

the deepest philosophical affinities with Quine's rejection of analyticity, the former still today retains the mark of a serious, lasting problem while the latter has largely come to be piously ignored or relegated to the domain of mere historical relevance, of the era before Kripke resurrected inflationary philosophy. Should historians of Soames' stripe get their way, the folks at Conservapedia will be replacing 'C.', not for 'C.E.' but for 'K.' That lends a whole new irony to Quine's curious class of K-truths.

Both Quine's arguments against analyticity and Goodman's 'grue' paradox should be seen as critiques of a particular philosophical view of empirical engagement. Both serve to reject the notion that the world presents itself to investigators in broken-up chunks ready piecemeal for cognition and regurgitation without modification into communicable linguistic items; the idea that we read facts off of the world and can repeat them in words, ready as the world gives them to us. Such a picture of empirical thought, taken up in the tradition from British empiricism, requires the world to have the structure we can cognize it to have independently of any cognitive input on our part; it also requires our cognitive capacities and our language to "fit" onto that already given structure. Although this understanding of empiricism is not at all attractive today; indeed, it hasn't been attractive since the end of the 18th Century, it's rejection is of a piece with the rejection of analyticity and with the 'grue' paradox, an insight I've never seen articulated.

We can look at the old empiricism as overlooking the dynamic character of the "directions of fit" both from the world to the language we use to describe and theorize about it as well as from language to the world. It overlooks the fact that the

terms we use in our empirical, theoretical engagement with the world develop and modify their meanings *as a consequence of* that engagement. Likewise, our very understanding of the world, our constantly developing body of theoretical truths, is given in the terms of our theoretical terminology.

Let me briefly review Goodman's 'grue' paradox to show how this relates. Goodman is considering a formal account of inductive reasoning, which has it that inductive reasoning is inverse deduction. To clear up some minor problems, we can restrict this to inverse universal instantiation. Applying 'grue', which denotes green things observed before t and blue things not so observed, we get a troubling outcome. Although all emeralds so far (in the time before t) observed have been green, these emeralds (*because* they have been green) have also been grue. Thus, by inverse universal instantiation, the observation of all emeralds heretofore confirms both that (H1) all emeralds are green and that (H2) all emeralds are grue. The fact that evidence confirms multiple hypotheses might be acceptable except for the fact that H1 and H2 conjunctively justify inconsistent predictions, such as that the first emerald first observed after t will be both green and blue.

Given this unacceptable outcome, the "new riddle" of induction is how to figure out what makes 'grue' a bad candidate for inductive generalization and 'green' and 'blue' good ones. Goodman calls this distinction one of "projectability"; projectable predicates are just those whose instances are inductively confirming. A common and commonsense suggestion for drawing this distinction is to say that 'green' and 'blue' are basic, while 'grue' is built up out of the other two. Goodman's answer is that this "basicness" is not at all necessary. Were we in fact lacking the

basic terms 'blue' and 'green' but had basic terms 'grue' and 'bleen'¹¹, we would define the non-basic 'blue' and 'green' in terms of 'grue' and 'bleen'.

This answer has it that although 'green' and 'blue' are basic in this sense, they need not have been. However, to imagine speakers for whom 'grue' and 'bleen' are basic terms and 'green' and 'blue' are more complex concepts built up from them, is to imagine either a cognitive being of a rather different sort from ourselves or to imagine a world rather different from our own. Beings whose color perception were somehow indexed to their phenomenal experience of time might develop such terminology in their basic observational vocabulary. Or, we might have developed them in our basic observational vocabulary were it the case on a broad scale that new tokens of natural types often appear differently colored at traceable moments in time, say, in accord with the phases of the moon or something of that sort.

The lesson of this rejoinder to one answer to the 'grue' paradox is that the terminology we use to describe the world develops through our interactions with it. This is to deny one of the necessary "fittings" of classical empiricism. According to that view, the categories and classes of cognizable empirical significance are both already there in the world, prior to our investigative engagement with it and irrespective of that engagement. On the picture Goodman's view assumes, the contribution of the world and the contribution of language we use to describe it in our theoretical offerings are developmentally conjoined in a way that suggests subtracting one from the other misrepresents both.

In this sense, Goodman's 'grue' paradox attacks classical empiricism at the

¹¹ In our terminology, 'bleen' means the same thing as 'grue' except with 'green' and 'blue' transposed.

same point that Quine's attack on analyticity does, except from what we might call the "opposite side". Goodman's argument attacks the notion that it makes sense to talk about empirically significant content in abstraction from the manner in which that content is stated and communicated. That is, he attacks the idea that the "word contribution" can be understood in isolation from the "world contribution". The response that 'blue' and 'green' are basic, if it is to be at all explanatory, relies on this problematic idea, because it suggests that basicness is a prerequisite of empirical inquiry (that is "read off" the world) rather than being an interactive consequence of such inquiry. Where Goodman's attack is on the idea that the "world contribution" of facts can be isolated from the "word contribution", Quine's attack on analyticity focuses on the same divorce from another angle, the idea that the "word contribution" – meaning – could be understood in isolation from the "world contribution" – contingent facts. Thus, the characterization of analytic statements is as "true in virtue of meaning alone" and "irrespective of what the facts happen to be".

In the manner just suggested, we can see the rejection of analyticity and Goodman's 'grue' paradox as similar critical responses directed from opposite ends of the traditional empiricist conception of the way language relates to the world. That conception has it that the world provides ready-made facts, which can be ascertained and reported in language without language influencing the facts or facts influencing the language. Giving up this picture requires giving up the notion that there are domains of language which are not influenced by facts, or – giving up analyticity; and it requires giving up the notion that there are "raw" facts unaffected

and irrespective of the language in which we report them, or giving up the notion that empirical observation (and generalization) is not influenced by the relative entrenchment of linguistic categories. The 'grue' paradox, and the view of observation it indicates, understood in the way I have suggested, is neither a consequence nor an implicating premise of Quine's (or anyone else's) rejection of analyticity; it would be *possible* to accept the paradox as genuine without accepting the rejection of analyticity. However, I've introduced and explained the relation between the two precisely because the possibility is a narrow and an odd one. The more general idea is that true sentences intermingle language and the world in a way such that neither language nor the world can be "subtracted" out of the other without misconstruing both. And, while it's not at all fashionable to accept the prevailing arguments against analyticity (even, in some egregious cases, like George Bealer's *Routledge Encyclopedia of Philosophy* entry on Analyticity, the fact that few philosophers accept Quine's argument is itself taken to be an argument against it), virtually no one rejects the idea that the 'grue' paradox presents a genuine problem. It is, therefore, a strange outcome that Quine and Putnam's arguments are generally rejected and Goodman's 'grue' problem generally taken seriously, while they are complementary consequences of the same radical rejection of a fundamental tenet of classical empiricism. This strange refusal to couple may indicate that many philosophers do not understand the deeper implications of Goodman's paradox, or that they fail to recognize the common lineage between it and the rejection of analyticity, or both; it may also be due to the fact that the thesis (especially Quine's version) challenges the viability of traditional philosophical methodology in ways

that are immediately obvious; though, I would argue, Goodman's paradox, less obviously, challenges that methodology on exactly the same grounds.

It may be replied to this argument, however, that although most philosophers take the 'grue' paradox seriously, few accept Goodman's theory of projection as a way of "solving" the paradox. And, moreover, it is the theory of projection (not necessarily the paradox itself) that informs his retort to the attempted answer that 'green' and 'blue' are more basic than 'grue' and 'bleen'. This is fair; the riddle is more popular among philosophers than is Goodman's solution. However, while Goodman's pragmatic solution, formulated in his theory of projection, may fail to satisfy many, workable options that do not incorporate the spirit of Goodman's pragmatic solution have not appeared, even at the most sophisticated levels of contemporary research on induction and confirmation theory. For example, Patrick Maher has developed an ingenious solution that explicates projectability in a formal system as "projectable across another given predicate" (Maher 2004). However, since the distinction Maher develops is a formal one, it wholly lacks the resources to identify the predicates we would presystematically identify as projectible and distinguish them from the non-projectable ones. Thus, a tempting and misleading way of describing Maher's achievement is to say that he's provided a way of formally describing the facts that 'green' is "projectible across 'observed before t '" and that 'grue' is not "projectible across 'observed before t '". This way of describing it is misleading precisely because the distinction is a formal one and is given in terms of predicate schemata, not actual predicates. Insofar as the schemata can be said to have semantic content, that content is described purely in terms of other

predicate schemata, not in terms of predicates. So, for example, we might define ‘G’ in the following way: $\forall x(G'x \leftrightarrow ((Gx \bullet Ox) \vee (\sim Gx \bullet \sim Ox)))$. Maher’s analysis of “projectability across a predicate” would render ‘G’ here projectible across ‘O’ and ‘G’ not projectible across ‘O’. The problem with this, and with the move to formalism as a response to Goodman’s problem generally, is that nothing in the account guarantees or even indicates that ‘G’ stands in for ‘green’ and ‘G’ stands in for ‘grue’. Thus, this account, without appealing to some sort of pragmatic condition relating the formal calculus back to our actual inductive practices, cannot distinguish for us the difference, in terms of projectability, between ‘green’ and ‘grue’.

D. Putnam and “Truth Come What May”

In much the way that the particular notion of analyticity that can be preserved by accepting Carnap’s language systems approach can do nothing at all to assist projects of traditional philosophical analysis, Hilary Putnam, in “The Analytic and the Synthetic” (Putnam 1975) rejects the philosophically salient variety of analytic statement while retaining a place for what he regards as an entirely banal and unilluminating variant of the distinction. In this section, I will present and critique Putnam’s arguments to this effect. Of the arguments directly posed about analyticity I’ve discussed in this chapter, Putnam’s view on the topic is the closest to being correct. I will argue, however, that his positive view has some flaws and that an argument similar to the one he uses regarding the non-analytic character of statements involving “law-cluster” terms can be appropriated and applied to his

own distinction between “law-cluster” and “one-criterion” terms. This appropriation casts doubt on the value of the distinction for illuminating any reliable difference between non-analytic statements and analytic ones.

Putnam’s view is that Quine is mostly right in his attack on analyticity but wrong that there are no analytic statements; he thinks there are no analytic statements in empirical science, but there are such statements in more ordinary contexts to which groups of exceptionless laws do not directly pertain. In order to underwrite these claims, Putnam draws on a distinction he develops between law-cluster and one-criterion words. Since Putnam agrees with Quine’s arguments where their purview is restricted to naturalistic contexts (and these are the only “first rate” contexts for Quine), I will just briefly sketch the gist of Putnam’s comments on law-cluster terms; then I will bring up some worries with even the very narrow notion of analyticity that Putnam argues can be preserved by his category of one-criterion words.

Throughout this chapter, I have discussed analyticity mostly under the idiom of “truth in virtue of meaning alone”. If analytic statements are true in virtue of meaning alone, they can also be characterized as “true come what may”, or true regardless of empirical fact. Thus, a change in truth would only come about as a consequence of a change in meaning of the terms and thereby a change in topic. Putnam focuses on the characterization of analytic statements as “true come what may” in his argument that there is no use for analyticity in science. In this, he relies on two primary examples, the principles of Euclidian geometry (1975: 46) and the Newtonian definition of kinetic energy as $e = \frac{1}{2}mv^2$ (42). These examples are

apposite because they both represent what were at one time foundational assumptions of physical inquiry, which were not open to rejection based on isolated experimentation; however, both were later rejected in light of both empirical observation *and* novel theoretical development. As aspects of empirical theory not open to rejection based on isolated experimentation, they look like they *might* be candidates for truths held “come what may”. However, taking the empirical significance of the body of physical theory as a whole, it is clear that the fact that neither Euclidian geometry nor the classical definition of ‘kinetic energy’ could be rejected by isolated experiment does not entail that they are held true come what *observations* may. The status of these special truths is simply as very entrenched such that it would take more than an aberrant observation to reject them – it would take the availability of a plausible alternative theory to replace them. And, furthermore, just this sort of thing is what actually happened. Euclidian geometry was rejected when alternative geometries had been worked out that more aptly conjoined with current understandings of optics. Newton’s definition of ‘kinetic energy’ was rejected when Einstein had developed the rival theory of relativistic kinetic energy.

This reconstruction of the scientific reasoning in both cases leaves open the important question why these rejections of foundational scientific principles don’t amount to changes in topic (and thus leaves open the possibility of classing them as analytic). To answer this, Putnam appeals to his concept of a law-cluster term, a term that appears in a network of related exceptionless laws in science. His idea is that a concept, such as *kinetic energy* appears in all sorts of scientific laws, not just in

Newton's definition or in Einstein's, and a change of the one law – from $e = \frac{1}{2}mv^2$ to $e = mc^2 + \frac{1}{2}mv^2 + \dots$ – doesn't entail changes in the other laws in which the term e appears. This is the sense upon which Putnam is drawing to designate the concept as a law-*cluster* concept; it appears in and takes its identity from a cluster of related laws such that some number of them could be changed without there being a change in topic. In this manner, Putnam can also be read as elucidating an aspect of Quine's holism, since he specifies particular ways in which the network of theoretical statements are related to each other. Also, other considerations about our ways of describing this sort of theoretical change can be brought to bear in support of Putnam, such as the fact that it is customary to say that Euclidian geometry is wrong, as a metric for space, or that Newton's definition of kinetic energy is correct only for systems of certain magnitudes. If changes in theory amounted to changes in topic in the way suggested, these ways of describing theoretical development would be mistaken; instead of saying that Newton or Euclid were wrong, we would have to just say they were talking about something else from what we're talking about when we use the same phonemes and inscriptions. Moreover, the sort of networked "law-cluster" character of scientific terminology can be offered as an answer *why* we talk about theoretical development in the way I've indicated and why we're right in doing so. That is, our grip on theoretical terminology, since it pervades various aspects of theory, can be seen as part of an evolving body so that we can modify some elements, even central ones, such as definitions, without changing the topic to something else; our use and understanding of a theoretical term relies on various laws, not just a single one.

Putnam's next move is to distinguish "one criterion" from "law-cluster" terms, and thereby retain a place for analytic statements outside scientific contexts. His idea is that, since the reason statements involving law-cluster terms cannot be analytic is because they appear in other exceptionless laws, terms applicable according to one criterion alone will not appear in other such laws. Changing the "one criterion" changes the topic, because we don't have a body of other theory to give the term meaning apart from the single criterion whereby we apply the term. Putnam's central example of a one-criterion word is 'bachelor'. He says,

This is not to say there are no laws underlying our use of the term 'bachelor'; there are laws underlying our use of any words whatsoever. But it is to say that there are no exceptionless laws of the form 'All bachelors are ...', except 'All bachelors are married', 'All bachelors are male', and consequences thereof (1975: 57).

Furthermore, he notes that 'bachelor' may appear in any number of statistical (non-exceptionless) laws, but maintains that these laws don't matter in the right way because

these cannot be incompatible with 'All bachelors are unmarried men'. For the truth of a statistical law, unlike that of a deterministic law, is not affected by slight modifications in the extension of the concept. The law 'ninety-nine per cent of all *As* are *Bs*', if true, remains true if we change the extension of the concept *A* by including a few more objects or excluding a few objects. Thus, making *slight* changes in the extension of the term 'bachelor' would not affect any statistical law about bachelors, but by exactly the same token, neither would *refusing* to make such changes (58).

Putnam entertains one objection to this argument and cedes its claim that just because we don't know of any exceptionless laws now pertaining to bachelors it could turn out that there are some, however unlikely this might seem from our present point of view. He concludes, however, that just because this is the case

doesn't mean that 'all bachelors are unmarried' isn't analytic, it's just the case that it need not remain analytic in all logically possible futures, such as futures where bachelors turn out to represent a natural kind of some sort.

This slight ceding of ground is fine. However, Putnam fails to entertain what, by my lights, is a more serious objection to his argument that one-criterion terms can be used in analytic statements. My objection to Putnam is that he hasn't motivated the reliance he places on exceptionless laws and in placing reliance on such laws in the way he does he holds non-scientific terminology up to scientific standards. If we look back at his arguments regarding law-cluster concepts, what is doing the work in that case is the fact that, for these scientific terms there are many, many laws other than the one being thrown out to which we can look to gain a grasp of the meaning of the concept. This isn't offered as a theory of meaning, but it does rely on the suggestion that whatever the meaning of a law-cluster term is, it can survive the rejection of a law involving that term by piggybacking on our understanding of other laws involving the term. By using the distinction between law-cluster and one criterion terms, he makes the point well that generalizations involving one-criterion terms *cannot do the same thing*; namely, they cannot survive a rejection of the single criterion on the back of other exceptionless laws, because no such laws are available. But this does nothing at all to answer why exceptionless laws should be needed to keep the original meaning of the term afloat - especially in non-scientific contexts. Why couldn't we keep our grip on the meaning of the term by relying on all of those statistical laws, or, more ordinarily, otherwise relied-on stereotyping generalizations about the term's referents? It seems to me that this is

more like the way that *most* of our words work. Who, besides perhaps philosophers, can readily name off a single exceptionless law about most of the substantives they use in ordinary speech? In most contexts at least *access* to exceptionless laws about the topics in question doesn't seem to be needed in order to be clear on the fact that we're talking about one thing instead of another.

In response to this, Putnam would probably say that he's not talking about the way most of our words work; he's talking about the way that one-criterion words work. On this picture, words in theoretical contexts appear in lots of exceptionless laws; words in mundane contexts appear in no such laws; and one-criterion words appear in exactly one such law. They, therefore, represent a kind of artificial middle between systematic and non-systematic discourse. I call this middle point "artificial", precisely because of the single exceptionless law that features the word; the idea is that were the type in question the right kind of thing to be in exceptionless laws in the first place – a natural kind, say – then it would be in lots of such laws, because the various statements of systematic theory are interrelated. A response of this type just serves to refocus the original question, however, that it's not clear why it is exceptionless laws (instead of statistical ones or stereotypes) that matter for one-criterion words. If non-systematic terms for which there are no exceptionless laws can survive changes in general beliefs about the type in question – indeed, if they can have a stable meaning at all, it's not clear why one-criterion words should need a background of exceptionless laws to keep their reference straight under the rejection of the one-criterion. In defense of his distinction, here Putnam would offer his third of four conditions for analyticity: "the [one] criterion is

the only one that is generally accepted and employed in connection with the term” (1975: 65). There may be lots of connotations to the term and it may appear in general and stereotypical contexts, but one criterion, the one exceptionless law is generally used as the single determining factor for whether the word applies or not. So, for example, we might connote bachelorhood with all sorts of things, such as a messy apartment, sexual profligacy, and “commitment issues”, but we only really determine whether someone is a bachelor by determining whether they are male, adult and unmarried.

In just this sense, the distinction Putnam advocates between analytic statements and others draws to some degree on actual facts of usage. It seems the fact that a term has just a single exceptionless law pertaining to it doesn’t by itself underwrite the notion that the statement of that law is analytic; rather it must also be the case that we treat that law in particular linguistic and epistemic ways, namely, that we identify tokens of the type in question primarily according to that law and that we justify our claims about such tokenings by appeal to that law. Thus, analyticity comes down, at least in some sense to how we use sentences that are candidates for analyticity. And this is reflected in Putnam’s ultimate justification for allowing analyticity in what he determines to be unimportant cases; he allows it because, as he says, “it can do no harm” (1975: 59).

When Putnam says that allowance of analyticity in cases where law-cluster terms are not involved “can do no harm”, he means, in a relatively narrow sense, that it can do no epistemic harm. In the sense he has in mind, he is correct, it cannot do the kind of harm that regarding sentences involving law-cluster terms as “true

come what may” could do; it cannot undermine the progress of systematic empirical inquiry by forcing us to conclude that we’ve changed topics every time a new theoretical development is achieved. But it can do harm morally, and, if ethics admits of truth and falsity, then it can do harm epistemically as well. It is puzzling that Putnam chooses ‘bachelor’, which he understands as synonymous with ‘unmarried adult male’ as an example of a term whose definition is innocuous and one which is disjointed from outside considerations that might issue in modifications to its meaning were it not insulated by the one exceptionless criterion whereby we understand it. The very idea that this or any other term loaded down with content dichotomizing both gender and sexual orientation is innocent in the way the Putnam seems to think is ethically naïve.

Ways in which Putnam’s analysis “can do harm” in the case of bachelor are at least two. It can be used to assert bachelorhood of those who are not bachelors, such as W. H. Auden, during his long-term partnership with Chester Kallman. Auden was an unmarried man, but despite that fact he hardly fits the ordinary concept of bachelor. If we hold strictly to Putnam’s view that bachelors are all those who are unmarried men, regardless of the other connotations of the term ‘bachelor’, we will misconstrue important facts about Auden. An expectable response to this is likely to point out the fact that marriage was not even a legal option for Auden and that had he been able to marry, he would have, especially since he considered his relationship with Kallman to have been a marriage. Perhaps he would have married, perhaps not. At any rate, the fact that he might not have and the fact that any number of similarly significant relationships, both heterosexual and homosexual,

involves no pretext to marriage undermines whatever significance this point might have. Given these facts, one way that holding fast to a law that equates 'bachelors' with 'unmarried men' can "do harm" is by underwriting claims about the status of people's private lives that they would not themselves accept, such as that Auden or Kallman is a bachelor. This is moral harm enough, I should think, to reject Putnam's analysis of 'bachelor'.

The outcomes of Putnam's view aren't much better for women than they are for those with life-partners who cannot or don't wish to marry. 'Bachelor', in contemporary usage, connotes a free-wheeling lifestyle and does so with a light-hearted and playful sense of suspicion. Bachelors are often thought of as playboys, and while they may lack a certain sort of seriousness, their detachedness is taken to be enviably understandable. "Bachelorettes"¹², however, stereotype into two classes: spinsters and sluts, the first of which fail to be desirable to men and the second of which fail to control desires of their own. Such women are unwed precisely because only a desperate fool would marry them, while bachelors are unwed because it's so much fun to carouse without commitment. Thus, while we could follow Putnam's lead with 'bachelor' and understand 'All bachelorettes are unmarried women' we thereby advocate the use of standards that variously connote ethical bias that cuts along gender lines. This also seems to be a moral reason why following Putnam's account "can do harm".

About the first issue I raised regarding non-married, committed people, I am

¹² 'Bachelorette' has become less prevalent, except for in special contexts, e.g., 'bachelorette party'. Compare this with 'poetess', which is not used at all in contemporary speech. 'Actress' is also becoming more rare, although it is still used in some influential contexts, such as the Academy of Motion Picture Arts and Sciences.

not aware of any strong counterargument that Putnam can make. He might try to push off the problem onto 'marriage', claiming that the couples in question are actually married, and the real culprit is too narrow a definition of marriage, not too wide a definition of 'bachelor'. A cursory acquaintance of current legal debates concerning same-sex marriage reveals this way out as stickier than the problem it is trying to patch. It also introduces an actual case that should make us skeptical about Putnam's third condition for analyticity that I cited above, that "the [one] criterion is the only one that is generally accepted and employed in connection with the term" (1975: 65). One of the arguments from opponents of same-sex marriage is that it violates the "definition of marriage". This is exactly the sort of argument that fits well into Putnam's picture; the underlying idea is that part of the definition of 'marriage' is that it, to regurgitate a hackneyed phrase, "is between a man and a woman". In defense of his view, Putnam could cite condition three, saying that this aspect of the definition is not "generally accepted and employed in connection with the term" 'marriage'. Maybe, maybe not. The fact that this is a contingent matter, dependent mostly upon the level of bigotry in a linguistic community doesn't bode well as a motivation for the condition. It leaves homosexuals in a community that disregards their rights without "a course of intellectual self-defense to protect themselves from manipulation and control", to borrow a well-drawn phrase from Noam Chomsky (1989: viii). The point here is that many terms which might fall under Putnam's heading of one-criterion terms have important political and ethical content and consequences, and the bottom justification that holding exceptionless generalizations involving these terms immune from revision is that "it can do no

harm". Such terms *can* do harm, of course, and reliance on the general rung of opinion to stand as the final arbiter of truth in such matters is certain to exclude many and to relieve them of grounds of rational self-defense.

About the second of my arguments, involving women, Putnam has a little more wiggle room. On this point, he can say that the harm done is, unlike with respect to same-sex and other non-married partners, done by the *connotations* of the terms, 'bachelor' and 'bachelorette', not by the denotations identified in terms of the single criteria, 'unmarried man' and 'unmarried woman', respectively. This is a reasonable response. However, a response is apparent that nevertheless tells against the claim that holding 'bachelor' immune from revision can "do no harm". The fact that there *are* two terms in the first place, a fact made necessary by the gendering of 'bachelor', contributes to there being different connotations that cut along gender lines. If 'bachelor' weren't for men only, then whatever connotations it had could just as well apply to women, and similarly with 'bachelorette'. So while the harm done to women by Putnam's version of 'bachelor' may not be strictly in virtue of its "one-criterion", this single criterion does partially provide the grounds for such harm.

My foregoing arguments have to do with factors specific to the case of 'bachelor' and 'All bachelor's are unmarried men', and therefore it does not necessarily generalize to other examples. Also, in his own defense, Putnam might point out that his article was first published in 1962, a time when the sorts of concerns with gender, sexual orientation and the philosophical status of marriage weren't as pervasively recognized. (Although, I should say, one of the nice things

about the Auden example is that Auden and Kallman's "marriage", predates "The Analytic and the Synthetic" by more than twenty years.) Still, since my argument focuses on specific ways in which the "harm" is done by maintaining a refusal to revise the specific statement 'All bachelors are unmarried men', it leaves open the possibility that there are other analytic statements that fit Putnam's view but which don't actually cause harm and for which it's not clear how they could. An example he mentions that might fit this description is 'All vixens are foxes.' Although, since 'vixen' is a biological classification, it's not clear why it shouldn't be considered like any other term in scientific discourse as open to revision upon the further development of theory. It might, for example, turn out that 'vixen', initially applied only to female foxes, is actually more theoretically useful as a significant biological kind of some sort. Perhaps some interesting biological role is usually played by female foxes but, under evolutionary pressure, some male foxes get ahead by filling that role, or some member of a related species exploits foxes by playing that role. It could also turn out that most female foxes have some genetic inheritance in common with a few non-females or non-foxes and that this class shares some noteworthy phenotypic characteristics. This sort of development is something like the actual change in biological taxonomy that classes whales together with hippopotami, as Cetartiodactyla, far more closely than it classes whales with sharks, although by phenotypic standards the latter pairing is far more obvious than the former.

So let me take stock of my discussion of Putnam's position before making my concluding remarks in this chapter. Putnam takes up the spirit of Quine's holism and

argues using salient examples that epistemically significant statements, such as those that play central roles in physical theory, are never analytic, in the sense of being “held true come what may”, despite the fact that they may not be overthrown as a consequence of isolated experimentation. He argues, however, that more trivial sorts of statements, statements that do not play a role in scientific enquiry, could be analytic, and he lays out necessary conditions for the ascription of analyticity in such cases. The central condition that distinguishes the trivially analytic cases from the deeply entrenched truths of scientific enquiry is that the identified term in the trivially analytic statements denotes its referents according to a single exceptionless criterion. Both Putnam’s negative case (against analyticity in science) and his positive case (for trivial analyticity) appeal ultimately to the actual practice of scientists and ordinary speakers. This sober methodology is a laudable one in philosophy; however, it is also underdetermining. With regard to the history of science, as Putnam appeals to it, this does not create problems; that is, it’s a fairly determinate question whether Euclidian geometry was given up as a fundamental metric for space or whether the Newtonian definition of kinetic energy is endorsed by physicists today. What is much less determinate or clearly answered by observation of speech behaviors is what the “actual linguistic practices” of English speakers with regard to the term ‘bachelor’ are. Given the sorts of gender and sexual orientation questions to which ‘bachelor’ and its connotations are sensitive (or, I should say, *insensitive*), there are bound to be speakers by whom ‘all bachelors are unmarried men’ is not held true. And, importantly, at least some of such speakers will be able to give reasonable arguments why they reject the “law”, and these

arguments (such as those I have given above) will be irrespective of ‘bachelor’'s failure to appear in a cluster of exceptionless laws. This becomes an issue of whose “actual linguistic practices” count and shows that appeal to usage, while worthwhile in the abstract, can fail on its own to deliver decisive insight, philosophical or trivial.

In some sense, these arguments leave Putnam’s view possibly intact, since they attack his central example more than they do his criteria. That is, there might still remain a space for analytic statements in Putnam’s sense; I have argued not that his criteria for identifying analytic statements are wrong but rather that they are difficult to apply and that Putnam himself hasn’t applied them well. Where Putnam sees ‘All bachelors are unmarried men’ as so philosophically trivial as to be unable to “cause harm” if “held true come what may”, I have suggested ways in which this holding true can and does cause harm – moral and political harm. In this way, the heart of my objection to Putnam is not that he’s wrong about analyticity, but that he’s wrong about what’s philosophically trivial.

This outcome reiterates a theme I have touched on variously throughout this chapter: Quine’s dogmatic naturalism. In Putnam’s division between one-criterion and law-cluster terms, we see a much less noxious but not altogether benign version of that dogmatism. Putnam does not want to say, with Quine, that the only true or “first rate” or meaningful claims are those within science; thus, he takes seriously the truth of claims like ‘All bachelors are unmarried men’. However, in the account of analyticity, he holds ‘bachelor’ to a scientific standard: the reason our grip on ‘bachelor’ *couldn’t survive* the rejection of ‘All bachelors are unmarried men’ is precisely because ‘bachelor’ sort of behaves like a scientific term – it’s got an

exceptionless law! – but it doesn't quite pass muster as such a term – alas, it's only got *one* exceptionless law. The problem with treating 'bachelor' in this way is that it rules out just the sort of points I raised above about women, committed gay men and bachelorhood, because it implies that bachelorhood no longer exists after the rejection of the "one-criterion".

E. Conclusion

In this chapter, I have considered a number of ways of making sense of analyticity. I have considered various possible formulations in terms of sameness of meaning, according to definition, semantical rules, and "truth come-what-may". In the cases of each of these attempts, explanatory satisfaction has been found wanting and positive answers to the hard philosophical questions have turned out to be either based on philosophical assumptions equally as questionable as analyticity itself or, in the case of Carnap, anti-philosophy, or, in the case of Putnam, outright philosophical error. And, what is more, although both of the philosophers I have commented on that accept some form of analyticity accept only an extremely deflationary version, profound problems nevertheless have been shown to arise even in the trivial cases. This bodes poorly for those philosophers who wish to use analyticity as a basic and non-empirical tool of philosophical trade, since insurmountable problems for justifying claims about analyticity in the case of 'All bachelors are unmarried men' are bound to return in spades for more philosophically robust topics.

Given these problems, I suggest we try another route in philosophy. Thus, in

the following chapter, I will detail explication as a philosophical methodology, which, unlike analysis, does not entail or issue in unexplanatory ascriptions to analytic truth.

CHAPTER 2

EXPLICATION

I. Explication and Prior Commitments

In the previous chapter, I tried to show that *conceptual analysis*, at least if that concept is understood with a certain degree of precision, fails as a methodology for philosophers. If conceptual analysis is underwritten by analytic truths, then it is useless for philosophy, because if there are analytic truths at all, they are trivial and fail to reveal anything insightful about philosophical topics. In this chapter, I recommend an alternative approach to what can be loosely referred to as analysis of concepts; this approach is explication; instead of carrying on with conceptual analysis in the original sense underwritten by analytic truths, philosophical concepts ought to be explicated. Explication, in the sense I recommend, is a process of inquiry whereby relatively vague, presystematic terms are replaced by more precise terms that preserve certain features of the prior term. Since explication does not require the new term to exactly match the antecedent usage of the original term, no synonymy relation, like those established¹³ by analytic statements, is required.

Explication, as a philosophical methodology, was first introduced and defended by Rudolf Carnap. Carnap's philosophical output is often identified with a rigid formalism and a preoccupation with scientific answers to philosophical questions. Indeed, sometimes the standards of scientific clarity and precision Carnap seems to demand of philosophy manage to seldom be met not only by philosophy but also by science itself. In this sense, the methodology of explication

¹³ Or, establishing, depending on the view.

particularly suits Carnap's tastes for the formal and systematic in logic, in philosophy, and in science. However, I suggest that we can recognize insufficiency in ordinary concepts without placing formalistic or systematically empirical constraints on satisfactory explanations in philosophy (or science, for that matter, especially if we include the social sciences as legitimate domains of inquiry). One tactic to cut along this pass would be to take Carnap as a starting point for looking for alternatives to traditional analysis, but to recommend a version of explication that is less restricted than the approach he recommends. As will become clear from textual evidence, however, this tactic is unnecessary – the portrayal of Carnap that leaves his methodology open to the kinds of objections that issue in the charges of formalism and scientism draw from misreadings by his critics, not from his own views, as stated. My recommendation (and Carnap's, it turns out) is comparatively minimal; instead of applying prior commitments, like formalism and empirical systematicity, to meaningful expressions and explanations, which underwrite the departure from ordinary usage, explication in my sense is guided by a basic pragmatics of what is explanatorily useful. In this sense, the core idea of explication as a philosophical approach is that philosophers should be explicitly willing to refine concepts from ordinary language, to make our concepts more precise, and to tolerate mismatches in use between our new explicated concepts and the ordinary concepts they refine and replace. In practice, this will mean, among other things, that some assertions that, for example, a consequence of an analysis is “counterintuitive” need not be taken as evidence against the rightness of that analysis. Understood in this way, explication is a method already employed by most

philosophers, and the point of advocating it explicitly is, thus, not to make a radical suggestion, but rather to lay bear certain epistemic norms already in place and to throw into high relief the differences between explication and analysis in the sense underwritten by appeal to analytic statements.

My defense of Carnap's method (understood as I have sketched) will follow in two sections. In the first of these sections, I'll detail the method using citations from Carnap, and I will discuss and dispatch the most prevalent objections to the approach. As I have suggested, these objections tend to amount to straw man arguments in that they attack features of some of Carnap's work, which are extraneous to explication as a methodology. In the second of the two sections, I will discuss a number of clear examples of explication, including Quine's use of explication in what he points to as a "philosophical paradigm" – his account of the ordered pair (1960: 257–261). This discussion will reintroduce the topic of prior or accompanying philosophical commitments that we will have seen influence reception of Carnap's method. The commitments that motivate Quine's appreciation of the method in the ordered pair case are orthogonal to Carnap's – instead of empirical systematicity and formalism, Quine is motivated by a taste for ontological sparseness. We need not share this taste in order to use explication. I will also discuss examples of explication, from the history of timekeeping and from economics that have no overlap with the formalistic or ontologically sparing tastes of Carnap and Quine. Through this, it should become clear that the central motivating factor of the method is its explanatory promise, whatever "explanatory promise" amounts to in particular cases. That is, departures from the satisfaction

conditions of ordinary terms and concepts are licensed by the sorts of inroads to inquiry and understanding made possible by these departures. Explication, then, doesn't simply help us avoid foregoing analysis due to the counterexamples that are liable to appear to any account of an ordinary language concept, its departures from these concepts create new lines of inquiry by expanding the set of quality theoretical tools at our disposal.

II. Carnapian Explication

Carnap describes his method of explication in the beginning of *Logical Foundations of Probability* (1962: 3):

By the procedure of *explication* we mean the transformation of an inexact, prescientific concept, the *explicandum*, into a new exact concept, the *explicatum*. Although the *explicandum* cannot be given in exact terms, it should be made as clear as possible by informal explanations and examples. The task of explication consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. We call the given concept (or the term used for it) the *explicandum*, and the exact concept proposed to take the place of the first (or the term proposed for it) the *explicatum*. The *explicandum* may belong to everyday language or to a previous stage in the development of scientific language. The *explicatum* must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific either logicomathematical or empirical concepts.

A number of objections have been raised to the method Carnap describes here. The main line of argument against explication has to do with the relation between *explicata* and *explicanda*. Here's a summary. Either an *explicatum* sufficiently matches the *explicandum*, or it doesn't. If it does match, then the *explicandum* was a sufficient term or concept all along and didn't need replacing. If it doesn't match, then whatever problems, stated in our ordinary language (using the *explicandum*)

would fail to be solved or clarified by explanations using the explicated term (the *explicatum*), because a topic change has taken place in the shift from ordinary to systematic language. The first horn of the dilemma is a relative of the “paradox of analysis”, going back to *Meno*, and having more recent formulations by Moore (1963) and Sosa (1983). Carnap’s notion of explication embraces the second horn of the dilemma, so I will pass by the worry about explications being unnecessary or uninformative.

An objection of the second sort has been made by Strawson (1963) and answered by Carnap directly (1963). Patrick Maher has argued that Carnap’s answer fails to meet the challenge made by Strawson and provides an answer of his own to the objection (2007). Maher’s defense of Carnap relies on a concept that Maher does not explain – the concept of an *explicatum* “corresponding well” with an *explicandum*. A detractor might argue that, since the notion of “corresponding well” between *explicata* and *explicanda* has not itself been explicated, it is not available – for Carnapian reasons – as a grounding distinction between those *explicata* that run afoul of the topic-change problem and those that don’t. Thus, it remains unclear how effective a tool “corresponding well” can be for responding to Strawson, since it seems open to variant interpretations, and is likely to draw the charge of question-begging. Looking at this convergence of arguments within a broader context, I think it should become clear that much of how Carnap’s work is generally characterized fails to correctly construe his actual views; for this reason, the general line of critique marshaled by Strawson largely misses its target, insofar as its target is Carnap’s actual position. Additionally, Maher’s discussion of Strawson’s argument

and his defense of Carnap show signs of this misconstrual as well. In order to clarify these issues, I will reconstruct and comment on Strawson, Carnap, and Maher, in turn.

Much of Carnap's work is devoted to the development of formal, systematic artificial languages and the project of "rational reconstruction" of whole divisions of discourse (scientific discourse, in particular). For this reason, his method of explication is wrongly equated with these projects. The correct way of understanding explication in relation to these projects is that explication is a general approach, of which formal system-building is a particular, and particularly specialized type. Clarity on the question of formalism is particularly important here. Carnap tends to get characterized as a philosopher who thinks all meaningful expressions must be made within the language of some formal system and that explication is a process whereby such systems are created; neither of these claims about Carnap is true, however. Formalism is not a necessary goal or a criterion for explication, nor does Carnap claim that meaningful assertions occur only within the context of a formalized language. It is true that Carnap's attraction to formalism and the formulation of language systems arises out of the same motivation as his commitment to explication – a motivation to make the criteria for assertion and disputation as explicit as possible – but this does not entail that Carnapian formalism is a constraint upon explication as a methodology. As will become clear from citations from Carnap below, explication is a process whereby concepts are made more explicit, and formal systems represent but one type of precise discourse; precision, however, does not always entail formalism. Furthermore, just as

formalism does not, as a criterion, constrain explication, neither does scientificity¹⁴. Just as Carnap is attracted to formalism in contexts where he finds it apposite, he's also interested in science, because science tends to incorporate rigorous standards for admissible explanations. But the latter does not entail that he accepts as meaningful only explanations that are scientific (any more than he accepts as legitimate only claims made in a formal language system). Given these clarifications, the line of argument pursued by Strawson becomes irrelevant.

Strawson's argument against Carnapian explication is that, as a methodology for philosophy, it is insufficient, because, instead of resolving the puzzles that arise in ordinary language directly, it changes the topic by replacing the ordinary terminology with formally defined or scientific terminology. The original problems are then sorted out within this new system of formal and scientific discourse. So, while the original questions were philosophical, the answers that this method can provide are scientific ones, which might be somehow related, but are nevertheless different. He says:

For however much or little the constructionist technique is the right means of getting an idea into shape for use in the formal or empirical sciences, it seems *prima facie* evident that to offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant—is sheer misunderstanding, like offering a text-book on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart (Strawson 1963: 504-505).

As I have maintained, Strawson is wrong to regard Carnap's method as necessarily formalistic or as replacing philosophical questions with scientific ones. Carnap

¹⁴ Whatever *that* is.

himself clarifies this misconception in his response to Strawson, and I will provide citations to this effect below. Before doing so, however, I'd like to address the argument directly, appealing to the operators of first order logic as examples.

I choose the logical operators for this discussion, because Strawson explicitly cites them as examples of Carnapian construction (1963: 503); a little investigation will show that, in these examples, the replacement of ordinary terminology with systematic terminology does not represent a change of topic in the way claimed by Strawson. Strawson regards the use of artificially-constructed systematic language to always fall within the domain of some scientific inquiry for which it has been constructed; he says such construction-making might get us to forget about philosophy and do intra-systematic work, but that work is science, not philosophy. However, when philosophers use the systematic terminology of elementary logic in order to clarify philosophical arguments — to point out instances of "affirming the consequent", for example — these uses are philosophical uses. Someone may affirm the consequent in a philosophical argument, because the ordinary terminology used in conditional statements ('if', 'only if', 'just in case', *et c.*) is often used in unclear and inconsistent ways. When this lack of clarity becomes evident, then we appeal to replacing concepts, like the conditional and conjunction, which are precise and whose application conditions can be explained rigorously. Indeed, the philosophical discussion might turn to questions of implication and entailment at that point, and the rules for use of the systematic terminology of truth-functional logic would be used as tools in an effort to resolve the dispute.

Strawson is likely to reply to this that much of the philosophical work being done in a case like this is in the translation of the ordinary language to the systematic language of elementary logic, and it's not the systematic language itself that is doing the philosophical work. Just how remedial an error affirming the consequent would be on the part of anyone seriously engaged in philosophical inquiry undermines this reply from Strawson. Philosophers don't often make this error because they have generally incorporated the systematic concepts of the logical operators into their habits of speech and thought. This is not translation or interpretation, it is systematic *replacement* of vague and imprecise terminology with clearer and more precise terminology in contexts where clarity and precision are needed, and this replacement is made possible by past explications.

As I have shown, in an actual case where a constructed system of terms (accepted as such by Strawson) is used by philosophers doing philosophy, this use, contrary to his argument, does not change the topic or somehow fail to address the original philosophical concern. Carnap identifies part of what motivates this confusion when he points out that Strawson “sees a sharp separation, perhaps even a gap, between everyday concepts and scientific concepts. I see no sharp boundary line but a continuous transition” (1963: 934). Also, to get clear on the relation of explication to science and formal systems, he says the following:

The only essential requirement is that the explicatum be more precise than the explicandum; *it is unimportant to which part of the language it belongs*¹⁵. However, since exact concepts are more easily found in the scientific part of our language, it will often be useful to define the explicatum in this part. Furthermore, exactness and clarity are best achieved by a certain degree of systematization. Therefore, the

¹⁵ My emphasis.

explicatum usually belongs to a systematic conceptual framework. But the system may be of a rather elementary kind as, for instance, the system of numerical words in everyday language. The use of symbolic logic and a constructed language system with explicit syntactical and semantical rules is the most elaborate and most efficient method. For philosophical explications the use of this method is advisable only in special cases, but not generally (Carnap 1963: 936).

Given that this represents Carnap's stated view of explication, it is clear that it differs sharply from the characterization by Strawson that has him removing all philosophical questions to scientific domains or formalized languages. The only way that the charge of changing topic could be brought against the actual method advocated would be if adherence to all prior usage of philosophical terminology in philosophical discourse that any intentional precisification of that terminology under deliberative investigation would itself represent a change of topic. The demand for such adherence would be quite obviously absurd; it would demand of philosophers that they incorporate all prior uses of philosophical terms into any analysis of those terms' concepts, even when such uses are patently vague or are known to lead to contradictions.

In his paper defending explication, Patrick Maher also discusses Strawson's argument. He does this, however, without identifying Strawson's false assumptions about explication as a method.¹⁶ For this reason, in part, Maher wrongly accepts Strawson's argument as cogent. Maher quotes Strawson's argument about giving a physiology book to someone wanting to know about the workings of the heart and then says:

¹⁶ Though, oddly, he does discuss these assumptions later in another section of the paper not dealing with Strawson.

Carnap replied that explication can solve philosophical problems arising in ordinary language because it gives us improved new concepts that can serve the same purposes as the ordinary concepts that created the puzzles; the problems are solved by using the new language instead of ordinary language in the problematic contexts. (2007: 333)

That Maher characterizes Carnap's position in this way indicates that he has ceded important contextual ground to Strawson from the outset and in doing so failed, with Strawson, to correctly characterize Carnap's position. As I have indicated above, explication is not an enterprise properly confined to the creation of artificial languages; it sometimes involves such projects, and it sometimes doesn't. Thus, according to Carnap, we don't necessarily solve problems and puzzles by "using the new language instead of ordinary language"; often – indeed in philosophy generally – explication is of particular terminology within ordinary language. The fact that this is Carnap's position (not the position characterized by Strawson or by Maher) effectively deflects the force of Strawson's attack, which assumes Carnap's use of explicated terminology takes place within the context of non-ordinary languages and thus represent a different inquiry altogether from the philosophical quandaries originally giving rise to problems and puzzles.

Beyond this, Maher seems to misunderstand the response to Strawson given by Carnap, and, as a consequence, regards it as inadequate. Here is an excerpt from Carnap's reply to Strawson:

If we find that the pocketknife is too crude for a given purpose and creates defective products, we shall try to discover the cause of the failure, and then either use the knife more skillfully, or replace it for this special purpose by a more suitable tool, or even invent a new one. [Strawson's] thesis is like saying that by using a special tool we evade the problem of the correct use of the cruder tool. But would anyone

criticize the bacteriologist for using a microtome, and assert that he is evading the problem of correctly using the pocketknife? (1963: 934)

Maher replies that

Of course, nobody would criticize the bacteriologist, but that is because the bacteriologist's problem was not about the pocketknife. However, the relevant analogy for "one who seeks philosophical illumination of essential concepts of non-scientific discourse" is someone who seeks knowledge of the proper use of the pocketknife; Carnap has offered nothing to satisfy such a person. (2007: 333)

This reply seems to almost inexplicably fail to understand Carnap's metaphor. In this metaphor, the relevant analogy to "one who seeks philosophical illumination of essential concepts of non-scientific discourse" is not someone who seeks knowledge of the proper use of the pocketknife; the relevant analogy is someone who wants to cut something with precision, to understand why the pocketknife fails to facilitate such cutting, and to understand what kind of modification or replacement of the knife would be required in order to cut with the desired precision. You start using a pocketknife to cut something finely and you find out that it's too crude; it leaves surfaces uneven, it warps the material being cut and yields generally unsatisfactory results. Analogous claims are suggested about the use of unrefined concepts from ordinary language when philosophical precision is wanted out of them – the appearance of puzzles, problems and paradoxes. These kinds of problems – like the problems that would arise when a pocketknife were used to make precise cuts – are what give rise to philosophical investigation of ordinary concepts in the first place. In order to avoid these outcomes, and thereby resolve philosophical confusions, Carnap suggests we refine our linguistic tools, just like we might refine our cutting tools if we want to cut with precision. Just as Strawson (along with most other

commentators) is sent off the rails in terms of understanding Carnap by the latter's affection for science, it seems the mention of the bacteriologist and the microtome has led Maher to misread the metaphor as suggesting two different goals, an "ordinary" one – understanding the pocketknife and a scientific one (though it's not clear what this one is exactly, just that it isn't "about the pocketknife"). But the goals of the pocketknife-wielder and the bacteriologist are the same: they both want to cut something. If we can extend the metaphor, it illuminates Carnap's view that ordinary and systematic languages exist in a continuous transition from the commonly vague to the explicitly precise. I take it that assertoric language, generally, is used to "cut up" the world into conceptually manageable parts. Under investigation we tend to discover that some of the cut parts don't fit together well; they've been cut in too crude a manner for certain purposes – their warped edges prevent their use in making precise claims about reality. In order to make these cuts more precisely, we might need finer cutting tools.

To summarize, then, Strawson's problem was that answers involving explicated terminology fail to satisfy the original concerns that give rise to philosophical inquiry, because these concerns arise out of uses of ordinary – non-explicated concepts. But, from Carnap's perspective, part of understanding how and why these problems arise in the first place is in recognizing how the concepts at work in even asking the questions are inadequate, due to vagueness, imprecision, or the tendency to generate contradictions. Thus, it's not a refutation of the approach to say that refined concepts can't solve the problems that give rise to inquiry, because these problems arise when it's not even clear what, specifically, is being

asked. An example of this is Strawson's own person "who says (with a sigh) that he wished he understood the workings of the human heart" (1963: 504–505). It's agreed that a physiology textbook would not satisfy such a person, but he at least ought to glean some of the insight he desires by exposure to well-formulated and confirmed theories of human emotion, interpersonal attachment, and erotic love. Taking Strawson at his word suggests that if none of these theories rely centrally on the concepts *human heart* or perhaps *workings of the human heart*, they would, as a matter of principle, fail as approximate answers to this forlorn philosopher's longing for knowledge. Once we replace the hyperbole of Strawson's physiology book with the kind of thing that we might reasonably take the hypothetical subject to be asking for, it becomes clear that there is no reason *at all* to think that answers employing more precise and systematic concepts than those in which the original query was (loosely and figuratively) posed fail in principle and for that reason to answer the original concern.

This last line of thought brings to the fore the question of "aboutness" and its relation to explication. This is a hard problem and one that has a long pedigree in philosophy (Goodman 1961), especially the philosophy of language. The whole Fregean distinction between *Sinn* and *Bedeutung* that gave rise to modern philosophy of language can be seen as a contribution to this issue, and, given its scale, I hope to remain uncommitted to any particular view on it. However, it bears pointing out that Strawson's argument against Carnap seems to assume an obviously false position on "aboutness" – namely that a philosophical question and

an answer to that question cannot be *about* the same thing unless they are both presented in the same terminology.

This is where Maher comes back in; he identifies a notion (“corresponding well”) filling the role I have identified as “aboutness” as a means of answering Strawson. I mentioned above that Maher gives his argument against Carnap’s reply to Strawson within the context of a paper defending explication. So, though Maher thinks Strawson has an argument that is unanswered by Carnap’s pocketknife analogy, he also thinks this argument *can* be countered, and he offers a counterargument himself. Within the context I’ve developed up to this point for clarifying Carnap’s position, I think it is useful to go through the argument Maher presents as deflecting Strawson’s attack, because it nicely illustrates some of the assumptions that are commonly brought to understandings of explication, which are assumptions I’ve taken pains to jettison.

Maher asks us to suppose we want to find out whether a sentence *S* in ordinary language is true. Now suppose we explicate the terms of *S* and generate another sentence *S'* using these explicata. *S'* turns out to be true, but this Maher admits, following Strawson, does not answer our original question – whether *S* itself is true. Maher says, rather, that his procedure can assist us in determining the truth of *S* in three ways:

- (1) The attempt to formulate *S'* often shows that the original sentence *S* was ambiguous or incomplete and needs to be stated more carefully.
- (2) If the explicata appearing in *S'* are known to *correspond well*¹⁷ to their explicanda in other cases, that is a reason to think that they will correspond well in this case too, and hence to think that the truth value of *S* will be the same as that of *S'*.
- (3) We can translate the proof

¹⁷ The emphasis is mine.

or disproof of S' into a parallel argument about the corresponding explicanda and see if this seems to be sound; if so, we obtain a direct argument for or against S . (2007: 335)

I have already discussed how explication can legitimately help in the first aspect of the procedure described by Maher. Though, in so doing, I have pointed out that this is an effect of explication that reveals wrong assumptions underwriting Strawson's critique in the first place; Maher, on the other hand, offers it as part of a direct answer to Strawson. These assumptions can be summarized as an unwillingness to take the *replacement* aspect of explication at face value. That is to say, these critiques assume a perpetual retention of the *explicanda* – they are never really replaced by the *explicata*, rather they persist as the concepts in use in not only ordinary but even in philosophical contexts. The second and third aspects of the procedure Maher describes indicate that even in defending Carnap's view, he resists the replacement aspect of explication and replaces this aspect with his own notion of "corresponding well". In concluding this section, I'll discuss why "corresponding well" and the second and third aspects of Maher's procedure for judging this relation provide unsatisfactory substitutes for replacement.

Maher says "If the explicata appearing in S' are known to correspond well to their explicanda in other cases, that is a reason to think that they will correspond well in this case too, and hence to think that the truth value of S will be the same as that of S' " (335). This way of relating explications to puzzles that give rise to philosophical queries overlooks that explications may specifically depart from the norms of ordinary terminology they are designed to replace. That is, an *explicatum* may be designed to track only certain aspects of the use of the *explicandum* to which

it corresponds, and these aspects may be highly specialized or limited. Moreover, it may be a clear feature of an *explicatum* that it generates any number of pairs of sentences S and S' of Maher's type that differ in truth-value. For example, 'Unicorns have five legs' is, if anything, false; but it's explicated cousin, ' $\forall x((x \text{ is a unicorn}) \rightarrow (x \text{ has five legs}))$ ' is true. Maher doesn't say that "corresponding well" entails truth-value parity in every case, so he might reply that even examples where a vast number of truth-value mismatches obtain could still count as "corresponding well". This seems to be a narrow path for him, though, because, the notion of "corresponding well" is given as a reason to believe that the sentences in the pair share common truth-values; if well-corresponding terms give rise to large numbers of sentence-pairs whose members diverge in truth-value, then "corresponding well" couldn't be a very strong reason in favor of drawing the conclusion that S is true given that S' is.

Just as (2) in Maher's procedure for evaluating the relevance of explications constrains the explanatory value of explications by tethering them to ordinary terminology, so too does (3): "We can translate the proof or disproof of S' into a parallel argument about the corresponding explicanda and see if this seems to be sound; if so, we obtain a direct argument for or against S " (335). A central motivation for what we're doing when we explicate concepts is to try to improve on the functionality of our prior terminology; in some way, our prior terms are failing to provide us with insight into some aspect of reality, because they fail to track the world to a sufficiently precise or complex degree to satisfy our interest for understanding. This insufficiency, therefore, includes the explanatory norms we

deploy in assessing our own comprehension. Given this, there is no reason *at all* to think that (3) should be a constraint on the value of an explication. Our prior notions of what “seems to be sound” regarding arguments for and against *S* are precisely the reason we would begin a project of explicating terminology in *S* – we recognize the intuitions about what “seems to be sound” in such a case as unreliable and falling short of our own adopted commitments to norms of rational deliberation. Thus, it hardly makes sense to reflexively judge the validity of more systematic investigations according to our unreliable notions of what “seems to be sound” that such investigations are designed to replace.

III. Explications

A. Time for Change

It bears repeating here that the difference between ordinary terminology and explicated terminology is not one so much of kind but of degree, and, in an interesting sense, pedigree. We replace our prior concepts with more refined ones when it becomes apparent that the prior concepts fail to suit our current purposes. These purposes may change, they may themselves become more refined or specialized and they may be multiple at a single juncture. An illustrative example of explications in my sense and Carnap’s is the introduction of hours of the day as a way of explicating daytimes or replacing the more imprecise terms, e.g., ‘morning’, ‘daytime’, and ‘night’, or even the still less refined ‘early’ and ‘late’. Examples of this sort work nicely, first of all, because they meet the only real criterion for explications – that the new terminology must be more precise than the terminology

it replaces – but also because they exemplify the continual aspect of explication as a process. While the ‘hours’, ‘minutes’, and ‘seconds’ of the 24-hour clock replace vaguer terms whose satisfaction conditions have more extensive gaps and overlaps, these terms themselves have become subsequently submitted to explication by more empirically systematic terminology where greater precision was needed for various purposes.

The history of time measuring and marking concepts is itself a story of continual and ongoing explication efforts. The need to measure and mark time, when this includes long stretches of time is twofold; some device for measuring and labeling the times of day is needed (a clock) and some device for recording and projecting long stretches of past and future time is needed (a calendar). Given the terrestrial situatedness of human beings relative to other elements of the galaxy and solar system, attempts to create accurate calendars based on lunar phases may have arisen as early as thirty thousand years ago (Falk 2009: 20). The central problem with such calendars is that they fail to coordinate various interests that give rise to the need for precise time-duration concepts and devices for organizing them. Lunar months are useful in that they are fairly regular and they mark stretches of time that are conceptually manageable to human beings, given our memory capacity and so forth, but once we scale our interest in timekeeping to include relatively larger scale needs, this way of marking time needs refining. One such need that arises in part out of agricultural technology is the need to coordinate the change of seasons within the stretch of a year. The duration of moon phases does not coordinate well with the changing of terrestrial seasons; a twelve-lunar-month year is too short, causing the

seasons to slip later into successive years and a thirteen-lunar-month year is too long, causing seasons to occur too early in successive years. For this reason, the concept *year*, as understood presystematically as the span of four seasons proved to be too imprecise when conjoined with other precisely defined concepts, in this case the concept *month*, defined in terms of lunar phases.

Various local explications were introduced to deal with this mismatch; these involve different approaches to the adoption of an intercalative thirteenth month every several of otherwise twelve-month years in order to correct the drift of seasons. This was the case until the introduction by Julius Caesar of the solar calendar in 45 BCE. However, just as much earlier attempts to create usable calendars out of the ordinary concepts available proved to be untenable because these concepts failed to collaborate in ways that turned out to be consistent with their prior, separate uses, the same kind of mismatch occurs in the case of the Julian calendar, only on a much more modest scale. Where the approximation of agreement between the two relevant concepts in the transition from lunar to lunisolar calendars came to about one month's time annually, the approximation of disagreement in the case of the (solar) Julian calendar comes to about eleven minutes annually.

The major innovation of the Julian calendar was to detach the calendar month from the lunar month and add a leap day every four years, so that the twelve-month lunisolar calendar, which had 355 days and caused season-drift without an occasional intercalative month, would be replaced with a more accurate twelve-month solar calendar of 365.25 days (mean). In this way, the prior concept *month*,

which was understood in terms of lunar phases, was replaced by a more precisely defined concept relative to the solar calendar year, down to the number of days in each month. Just as the lunar/lunisolar transition involved coordinating alignment between timekeeping concepts of various scopes, so too does the transition from the Julian to the Gregorian calendars. Only in the latter transition, concepts coordinating with years are not *months*, but *hours*, and *minutes*. Well before the time of the Julian calendar's introduction, clocks of various varieties were in wide use; daytimes were recordable and trackable down to minutes. It turned out to be the case that the approximation of the solar year by the Julian calendar, scaled across not years but across numerous centuries, was too imprecise for some purposes relevant at the time. The actual solar year is about 365.2426 days long, a difference of 10 minutes 48 seconds from the Julian calendar's approximation of 365.25 annual days. This small discrepancy between the more exact measurement of the solar year and the concept of *year* controlling the civil calendar led to the calendar creeping ahead about three days every four hundred years. While it would take millennia for this slow creep to effect the kind of seasonal drift that was consequent of the lunar calendar, it did interfere with other, particularly contemporaneous purposes. By the middle of the second millennium, the Roman Catholic Church had been a dominant religion and political force for many centuries, and it had long considered Easter importantly coordinated with the spring equinox. This near eleven-minute excess of the Julian year meant that the dates of Easter and of the spring equinox would steadily drift apart. For this reason, Pope Gregory XIII introduced his reform in 1582, which corrected the discrepancy by restricting the incidence of leap years to

cancel out the additional three days every four centuries. Thus, “Every year that is exactly divisible by four is a leap year, except for years that are exactly divisible by 100; the centurial years that are exactly divisible by 400 are still leap years. For example, the year 1900 is not a leap year; the year 2000 is a leap year” (Lang 1999: 72). Notice that Gregory’s replacement of the Julian calendar’s concept *February* with a more strictly defined refinement of the concept does not simply lower the length of the calendar year by 10 minutes and 48 seconds and thus “correct” the approximation of the Julian calendar at all moments. Rather, it more closely approximates the calendar year to the solar year so that, for their purposes as far as they were aware of them at the time, it keeps the two concepts of *year* in agreement *closely enough* so that foreseeable problems don’t arise. In this way, the Gregorian explication can allow calendar years to slip ahead of their corresponding solar years by about 18 hours per century, but at the beginning of a century, it drops the calendar year behind the solar year by about 6 hours, by skipping the leap day. Over the course of the next century, the calendar year again gains 18 hours, making it just 12 hours ahead of the solar year by the end of the century, at which point a skipped leap year drops it behind the solar year by 12 hours. This continues until the succession of four centuries culminates in a century turn in which the century-long gain of 18 hours has brought the solar and the Gregorian calendar years into alignment. At this point, were the leap year skipped (as in other centuries not divisible by 4), the calendar would fall behind the solar year by 24-hours, so the leap year is not skipped, but is allowed to pass the solar year by 18 hours over the course of the century, restarting the cycle and thus preventing the slow but steady

acceleration of the civil calendar year past the solar year. As these details illustrate, the modification of the *year* concept as used by the calendar does not bring the civil calendar into perfect agreement with the solar calendar, but, as far as the objective that motivates it goes, the Gregorian innovation serves the intended purpose, because it keeps the civil calendar ahead or behind the solar year by at most 18 hours, and therefore keeps Easter within the same 24-hour span relative to the spring solstice.

The Gregorian calendar is still the dominant one in use today. However, where we have seen agricultural, cultural and religious purposes give rise to the need to supplant prior time-marking concepts in the past, more recently purposes of mass production, global telecommunications, navigation, advanced technology and competing branches of scientific inquiry have as well. It has been known since 1920 that the earth's rotation is not constant; it is steadily decelerating, and it also undergoes random changes in rotation duration as well as periodic ones (Nelson 2001: 511). As this became clear, it became important for some purposes to detach the concept of a fundamental unit of time, the *second*, from a system of time calculation determined by the earth's rotation. In 1967, the International Bureau of Weights and Measures explicated the term 'second' in terms of atomic radiation of a caesium atom, known to be regular with a far higher degree of precision than the earth's rotation. The 13th GCPM adopted the following definition: "The second is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom." And the 1997 meeting of the CIPM clarified that "This definition refers to a caesium atom

at rest at a temperature of 0 K" (BIPM). The length of the second as defined this way, however, is about the same as an astronomical second based on the mean solar day in about 1820. However, since the earth's rotation has been slowing about 1.4 ms per century, the day is now 2.5 ms longer than it was in 1820. This difference makes up about 1 s per year (Nelson 2001: 509). As a consequence of this mismatch between the more precisely regular atomic clock that regulates civil time and the civil calendar to which it is tied, organizations worldwide that monitor and track international clocks, such as the US Naval Observatory, must collectively insert leap seconds about once a year (but at irregular intervals) in order to keep world time-keeping systems in synchronicity and to keep them from moving behind the pace of the earth's rotation. The need to do this requires significant logistical effort and incurs great expense from many involved actors worldwide. From the standpoint of physics, which needs regularity in timekeeping above other features, there is little need to keep up this project of maintaining the leap second, and indeed for the public and business generally, the calendar's gradual creep behind the earth's rotation is sufficiently slow so that very little by way of disturbance of ordinary activities would occur were world time to be detached from terrestrial rotation altogether. As it stands, if we were to stop using leap seconds, our calendar would end up about an hour behind the earth's rotation – in 600 years, a future whose problems are scarcely imaginable to us.

However, from the standpoint of current astronomy, and the earth-bound observational perspective of terrestrial astronomers and their telescopes, the difference of one second per year in the calendar creates serious problems not after

centuries but after only a few years. Thus, regarding the concepts relevant in the calendar, this is a case where various purposes are at odds with one another and thus, there remains competition as to exactly which way is best to make those concepts precise.

B. A Change in Time?

Up to this point, I have discussed various explications of time concepts from the calendar and the clock with the aim to indicate how these explications involved formulating relatively more precise concepts than were previously available for antecedent purposes and thus these purposes were able to be served by the newly explicated concepts. It should be evident from the ongoing debate regarding the leap seconds and the atomic second *vis-à-vis* terrestrial motion and astronomical inquiry, however, that conceptual precision is not a characteristic to be valued always in relation to some purpose. Such purposes might be quite broad and even vague, such as the desire for “explanatory usefulness”; they might also be nefarious, as, for example, the Gregorian reformulation of ‘February’ must have been for many. Given this, I want to emphasize here that not all explications prove useful in inquiry. I have offered a deflationary account of explication, where the only criterion for a concept to be an explication is when it replaces a prior concept and is more precise than the concept it replaces. From this it does not follow that the explication therefore is necessarily going to be more useful than the prior concept at satisfying the investigative motives for which the prior concept was insufficient. That is, not all conceptual precision is accurate for relevant purposes. The atomic second is an

example here; had it been defined in terms of the (longer) solar second of 1967, the year of its adoption as the world standard, instead of the solar second of 1900, then the definition would have, for a time, been more accurate, not only for the purposes of physics but also for those of astronomers.

So explications can go awry. Modern economics features examples of this, the vast significance of which would be difficult to estimate. Modern economics has developed logistical models of economic behavior that have been impressively successful at describing and predicting certain sorts of abstract, complex interrelations between agents, such as is formalized in general equilibrium theory (Sen 1988: 8). However, these models (and modern economics generally) explicate certain key concepts in highly specious ways. Two assumptions standardly underwriting modern economic analysis are that human behavior is rational and that rational behavior is behavior that maximizes self-interest (10–12). Consequently, we can see the dominant strains of the discipline employing explications of the key ordinary concepts *human behavior* and *rationality*; within modern economics, human behavior is rational behavior and rationality is the maximization of self-interest. Clearly, to anyone not indoctrinated to the dominant dogma, the adoption of these explications legitimates patently false claims. But it is important to be clear that noting this is not enough to reject these explications out of hand. To clarify this, it will be helpful to draw on a discussion of explication by Quine.

In §53 of *Word and Object*, Quine discusses a case of explication, what he calls a “philosophical paradigm” (Quine 1960). He was concerned with maintaining the

ability to use the term ‘ordered pair’ or the expression ‘ $\langle x, y \rangle$ ’ without making *ad hoc* ontological commitments; specifically, the ordered pair treats two objects as one in a way more restricted than the way a pair of objects make a set. For, example, the set $\{1, 2\}$ is equivalent to the set $\{2, 1\}$, but the ordered pair $\langle 1, 2 \rangle$ is distinct from the ordered pair $\langle 2, 1 \rangle$. Since he already accepts sets as ontologically permissible, he identifies Wiener’s definition of the ordered pair in terms of sets as representing a paradigmatic example of philosophical explication. The definition is “ $\langle x, y \rangle$ is identified with the class $\{\{x\}, \{y, \Lambda\}\}$, whose members are just (a) the class $\{x\}$, whose sole member is x , and (b) the class $\{y, \Lambda\}$, whose sole members are y and the empty class”(Quine 1960: 258). In addition to Wiener’s, Quine also cites other definitions of the ordered pair that couch the concept within the apparatus of sets and of pure mathematics. Since these definitions, taken together, are mutually inconsistent, and, taken individually, allow us to make claims about ordered pairs which the prior, ordinary notion of the ordered pair does not legitimate, it is clear that in some ways these explications each depart from the ordinary concept they are individually designed to define. But these differences and points of mismatch do not affect any use to which the term ‘ordered pair’ might be put in an actual mathematical context. Thus, while the various explications variously fail to match antecedent usage of ‘ordered pair’, they do so in ways irrelevant to what was wanted out of having a concept of *ordered pair* in the first place; therefore, Quine designates these mismatches as “don’t cares”.

This review of Quine’s account bears on the modern economics concepts of *human behavior* and *rationality*, because, it might be claimed, like ‘ordered pair’, that

the explication of these concepts within modern economics might fail to match our antecedent usage of the terms 'human behavior' and 'rationality', but these mismatches fall among "don't cares", because they don't have to do with economic analysis, just as the mismatches between the various explications of 'ordered pair' and antecedent usage don't have to do with mathematics. This argument fails, however. Explicating human behavior as rational and rationality as maximizing self-interest fails empirically within economic reality. It is simply indisputable that human beings are motivated by all sorts of values other than self-interest (such as moral values), that they sometimes act on these motivations and that sometimes these actions fall within and effect economic reality. Here is a very ordinary example: by far, the most inexpensive and convenient way for me to buy books is on amazon.com; since that company cancelled its web hosting account with WikiLeaks at the behest of Joe Lieberman¹⁸, I have chosen, against my economic self-interest, to refrain from patronizing the site and have urged others to do so as well.

This is a very parochial example, though the mismatch between human behavior as understood within economics and human behavior as observable in the actual world has been used to found mainstream theory whose consequences have been disastrous for some, and may still prove disastrous for all of us if they cannot be supplanted. The central article of faith coming out of modern economic thought, as based on the explications I've discussed, is that of the utility of free trade: "under a system of perfectly free commerce, each country naturally devotes its capital and

¹⁸ "WikiLeaks website pulled by Amazon after US political pressure", UK Guardian.
<http://www.guardian.co.uk/media/2010/dec/01/wikileaks-website-cables-servers-amazon>

labour to such employments as are most beneficial to each" (Ricardo 1951: 133). While it might be merely bad inquiry were the ways in which this theoretical approach fails to reflect reality simply wrong but value neutral, the consequences are more serious. Prescriptive, normative notions such as *rationality* or what is "*beneficial to each*" country influence the judgments of policy makers as well as our own judgments about what is economically permissible, and what is unpleasantly inevitable. The horrors consequent of this view are too many to enumerate; they include the creation of the "third-world", global economic collapse in the 1930s and today, vast and increasing disparities in wealth worldwide as well as within the wealthiest countries, and the continuing destruction of the earth's ecosystems. Given these dire facts, it would be wise not to count the ways in which the economic explications of human motivation and behavior fail to match actual human motivation and behavior among the "don't cares."

C. Explications and Philosophy

I might appear to have left myself open to attack in discussing a range of examples from scientific discourses (physics, astronomy, economics), since I have argued vigorously in the previous section that the Carnapian method of explication is consistent with but does not entail the replacement of ordinary concepts by scientific ones. Such an attack, however, would focus too narrowly on this range of examples as *scientific*, and in so doing would miss the ways in which they are importantly philosophical. What I have tried to bring out especially in the examples discussed from economics is that, the formulation of relatively precise concepts to

describe and investigate aspects of reality does not guarantee that we are able to therefore more deeply or objectively or truthfully describe or understand reality. That is, *precision does not entail accuracy*. So, while the concepts in use may have a high degree of precision in tracking certain specialized classes of observed phenomena, this precision may be quite irrelevant in terms of satisfying the initial investigative curiosity and puzzlement that gives rise to inquiry in the first place. Such precision might, in fact, be highly misleading, due to the complex formulae and consequent descriptions it makes available. The questions of whether a certain explicated concept or set of concepts manages to help us to more effectively get at the kinds of insights we were looking for in the first place are, precisely, philosophical, not scientific ones. The fact that moments of time have been explicated in terms of atomic radiation or that human motivation has been explicated by orthodox economics as rational self-interest does not settle matters, philosophically, with respect to the nature of time nor of human motivation. Rather, in instances of this sort, philosophers should maintain a conceptual/theoretical two-way checkpoint station, maintaining the appropriate flow of traffic between ordinary and explicated concepts, between everyday thinking about the world (and the worldviews and policy decisions it culminates in) and the more systematic investigative disciplines. Philosophers should synthesize and interpret insights like those built into vying physical and astronomical concepts of time periods and ultimately usher these concepts back through the checkpoint station, bringing them into contact with more common discourse and improving the range of expression and thought within that discourse. This is just to say that some of what we have

learned from systematic inquiry could be usefully incorporated into related concepts in ordinary discourse. In cases like this, effort on the part of philosophers to refine the concept for use in ordinary contexts is recommended, especially when refining use in such contexts could have significant consequences for ethical thought, epistemic fitness, and public policy. Alternately, when salient aspects of conceptual thought from the ordinary side of the checkpoint fail to appear on the theoretical side, such as with the profound deconstruction of human reality by the explication of human motivation as "rational self-interest", philosophers should shuttle these salient aspects, missing from current theoretical offerings, across the checkpoint and into the more systematic discourse of economics. This kind of project will often involve the introduction by philosophers of explicated concepts that can be taken up in systematic inquiry, where such concepts are found to be either failing to track some important aspect(s) of the ordinary concept they replace or to be lacking a consistent explicated formulation altogether.

The remaining chapters of this dissertation will be devoted to illustrating a group of cases of this last sort. Within empirical cognitive psychology, a whole subdomain of study for the past 30 years has been devoted to understanding the cognition of similarity and analogy. While many of the findings within this broad research are interesting and have proliferated a vast literature, very little use can be made of it in terms of developing a consistent theory of similarity cognition or analogy cognition, because this research has been conducted in the absence of explications of its centrally controlling concepts: *similarity* and *analogy* (and their complements). That is, this research has been conducted without precisely

designated concepts – with explications for ‘similar’ and ‘analogous’. To use the checkpoint metaphor, the central concepts of these research programs have passed through the checkpoint without a philosopher checking them out, making sure they make sense and making sure they track some coherent set of salient aspects of the pretheoretical phenomena they denote. And, in lots of ways I will discuss, the concepts of *similarity* and *analogy* being used in the very mainstream of empirical psychology research *do not* make sense as precisely specified concepts for grounding systematic research. In an effort to correct this, I will offer my own explications of ‘similar’ and ‘analogous’ (and their complements). Then I will turn to discuss past research on these concepts within cognitive psychology and indicate how my explicated concepts can be used to reframe these research programs, to critique and repair flaws in current methodology and data analysis, and, perhaps most importantly, introduce new avenues for inquiry.

CHAPTER 3

SAMENESS, SIMILARITY AND COGNITIVE PSYCHOLOGY

I. The Same, Only Different¹⁹: an Explication of Sameness and Difference in Terms of the Denotation Relation

I have spent the previous two chapters comparing analysis and explication as philosophical methods and, in doing so, have rejected the legitimacy of analysis and have defended a deflated notion of explication that avoids the critiques of the method's detractors. Ultimately, however, my aim is not only to abstractly recommend explication as an alternative to analysis but shows its fruitfulness by putting it into practice myself. What I have in mind, principally, is an explication of the term 'analogy', which will be based on explications of 'sameness' and 'difference', or the relations 'being the same as' and 'being different from'.

I identify sameness and difference in terms of the denotation of particular predicates:

(Two items, a and b, are the same with respect to P_2) *just in case*
(a and b are both denoted by some predicate P_1 , and P_1 is denoted by some predicate P_2)
and
(Two items, a and b, are different with respect to P_2) *just in case*
(There is no predicate, P_1 denoted by P_2 that in turn denotes a and b)

So, to fill this out by an example, consider three items: Bronzino's *Allegory on Venus and Cupid*, the *Venus de Milo*, and Botticelli's *Birth of Venus*. The two paintings are both oils, and the sculpture is marble, so we want to be able to say that the paintings are the same with respect to medium and that they are both different from the sculpture in that same respect. So, at first blush, we might say that 'medium' denotes

¹⁹ Thanks to Steven Wagner for this title.

‘oil’, which in turn denotes the Bronzino and the Botticelli, and that no term denoted by ‘medium’ denotes all three artworks. But this won’t quite do, because while ‘medium’ *does* denote oil, ‘oil’ does not really denote the two paintings. It denotes things that are oil, such as the oil used to make the paintings, but not the paintings themselves. This problem forces us, when analyzing particular instances of sameness and difference to be precise in identifying exactly in what respect the items of comparison are the same. So, the paintings after all are not the same with respect to medium *per se*. That is, since the predicate ‘medium’, more explicitly, is ‘__ is a medium’, the paintings are not the same with respect to what is a medium; rather they are the same with respect to what medium they are. So, a little care is needed in choosing the predicates to be used in the analysis. I see two ways of capturing some of what is meant pre-systematically by this sameness relation. First, we can identify P_2 simply as ‘medium’, which denotes oil paint (‘Oil paint is a medium’ is true.), and ‘oil paint’ denotes the actual paint in each of the two paintings. The paint in the Botticelli and in the Bronzino are the same with respect to medium, or what medium they are. Second, we can identify P_2 as ‘artworks categorized by medium’, which denotes oil paintings; ‘oil painting’ denotes the Botticelli and the Bronzino, so the two are the same with respect to what artworks they are, categorized by medium.

Another example. The term ‘art’ denotes each one of the three pieces, and the term ‘type of thing’ or ‘category of object’ denotes art; so, according to my explication, the three are the same with respect to type or category of object; this is so, since art is a type of thing and they are all art. And so on with respect to

whatever else might denote art. So, if it's true that art is denoted by such lofty predicates as 'designed to edify' or 'being one of the signal achievements of humanity', then these items are also the same in these respects.

With this last example, a problem with the explication seems to appear, since it may well be true that art is one of the signal achievements of humanity, and thus that the predicate 'is one of the signal achievements of humanity' denotes art. Given the account as it stands, this looks like a problem, because it legitimates the claim that, say, Bronzino's painting and another painting, which is of mediocre quality, are the same with respect to being one of the signal achievements of humanity. Even if Bronzino's painting deserves this appellation, probably the mediocre one does not. In this particular instance, the problem arises because the class, art, has properties not all of its referents have. The predicate, P_1 , commonly denoting two objects can be denoted by a predicate, P_2 , which does not denote all of the objects denoted by P_1 . Another way of saying this is that the relation of denotation is non-transitive: P_2 may denote P_1 and P_1 denote a , but P_2 does not therefore necessarily denote a as well. So we have 'being one of the signal achievements of humanity' denoting art, and 'art' denoting the decent painting, but 'being one of the signal achievements of humanity' fails to denote the decent painting.

This presents a challenge then, because it seems to represent a case where the explication commits us to saying that, for example, a merely decent painting is the same as a great painting with respect to being one of the signal achievements of humanity. At first blush, one fix would be to restrict the definition to cases where denotation is transitive. Will requiring this type of transitivity work to undermine

the problem? It works in the case mentioned, but requiring it excludes lots of other cases that we would not want to exclude. In fact, restricting the explication in this way would significantly undermine its reach and relevance. For example, both the Bronzino and the Botticelli are rectangular, and rectangular is a shape, so 'rectangular' denotes both paintings and 'shape' denotes rectangular, but 'shape' does not denote the two paintings: while they *have* shapes they are not shapes themselves, so 'shape' does not denote them. In this case transitivity of the sort being discussed fails, but this is a case we would not be able to give up; were our explication of 'same' to fail to provide us with a way of saying that two rectangular paintings are the same with respect to shape, then its implications would hold little sway in more complex and controversial cases.

How to fix this problem, then? Looking back at the original, troubling example and comparing it to the shape example might bring out what's essential. In the troubling example, the problem was that two things could be art and art be one of the signal achievements of humanity without the two things being the same with respect to being one of the signal achievements of humanity. In the shape case, the paintings, both being rectangular are the same with respect to shape, given that 'shape' denotes rectangular. Notice the difference in the second-level predicates 'being one of the signal achievements of humanity' and 'shape'. The former includes the present participle, 'being', while the latter doesn't. If we change the syntax of the second predicate to match the first, we'd get 'being a shape'. Here's the locus of the problem: the two paintings are *not* the same with respect to being a shape, at least not in the same way that they're the same with respect to shape. That is, they're the

same with respect to being a shape – because *they're both not shapes*, but they're the same with respect to shape because they are both rectangular. Cutting out the 'being' solves the problem in the troublesome case as well, since both of the paintings, the Bronzino and the decent one, *are* the same with respect to one of the signal achievements of humanity, with respect to art they're the same, because they both are art.

But there is no established distinction in predicate logic that corresponds to the distinction I need. The full, regimented version of both 'being a shape' and 'shape' would be the same; namely, '___ is a shape'. This is important, because I have no way of distinguishing between how the paintings are the same with respect to being a shape (because neither is a shape) and how they are the same with respect to shape (because both are rectangular). If there is no difference between the second-level predicates in these comparisons, then my criterion couldn't distinguish between the two (clearly different) types of sameness. It seems that the difference that the two ways of stating the predicate engender must have something to do with the respects, or ways in which two objects can be understood as being the same or different. The one with respect to shape, or what the shape of the object is, and the other is with respect to being a shape, or whether the object *is a shape* or not. What I need here is a way of differentiating our statement of the denotation relations of the predicates and objects so that it's clear whether we're talking about the version that includes the progressive participle or the version that excludes it, but this cannot be done in terms of predicates only, because the same predicate is at work in both versions.

The objective can be met by specifying conditions on the relation of denotation between the second-level predicate and the object. In the case where we're talking about just 'shape', the fact that the paintings are not shapes is irrelevant to whether they are the same with respect to shape. But in the other case, where the predicate is read as 'being a shape', this relation *does* enter into the analysis in an important way. We have the two paintings, both rectangular, which (rectangular) is denoted by 'being a shape'. But, in saying that the two paintings are the same in the respect of 'being a shape' is just saying that they are not shapes, or that the predicate 'shape' does not denote either one of them. This introduces what may look like a new sort of, somewhat trivial sameness not covered by my initial explication. That is,

For any two items, a and b and any predicate P,
(a and b are not in the extension of P) *just in case*
(a and b are the same in not being Ps)

But this type of sameness can be covered by my explication with just a little predicate creativity. For any two things that are the same in being not in the extension of some predicate, it is possible to formulate some new predicate that denotes them in common and is in turn denoted by some other predicate, in respect of which the objects are the same. So, the two rectangular paintings, which are not themselves shapes, are denoted in common by 'is not a shape', and not being a shape is denoted by 'type of thing a thing is not'. These moves solve the problem. Since, with this, there is no need to differentiate between 'shape' and 'being a shape'; we can just use 'shape' all the time, as long as we analyze the predicate at the appropriate degree of removal from the object. The confusion was all in placing

‘rectangular’ between the objects and ‘being a shape’ in the analysis, when the sameness (in not being shapes) between the two objects is dependent directly upon the failure of the paintings to fall within the extension of ‘being a shape’, regardless of whether they are rectangular or not.

A final problem that I’ve seen to be on the horizon for this explication appears when the predicate denoting the compared objects in common collides with the predicate denoting that predicate, when we want to say that the common feature of the compared items is also the *respect* in which they are the same. So, for example, the three artworks are all denoted by the predicate ‘represents Venus’. But what predicate denotes representations of Venus, such that we can say that the three paintings are the same in that respect? What we want to say is that they are the same with respect to what they represent; namely, that they are the same with respect to representing Venus. So, ‘represents Venus’, in cases like this, is both the first and the second level predicate, or, both P_1 and P_2 . This brings out explicitly a problem with my formulation of the explication that has been below the surface up to this point.

The problem is that predicates do not denote other predicates in the sense that I’ve been maintaining; ‘shape’ does not denote ‘round’, rather it denotes round. ‘Round’, the predicate, is denoted by predicates like, ‘is a predicate’, ‘maps round things onto the truth-value true’, and ‘maps non-round things onto the truth-value false’. It is this distinction that prevents my first explication from admitting that the three artworks are the same in the respect of representing Venus, because the second-level predicate ‘represents Venus’ does not denote the first-level predicate

‘represents Venus’. Or, which is another way of saying the same thing, the sentence “‘represents Venus’ represents Venus’ is false. The subject of this sentence does *not* represent Venus; rather, it denotes things that do. But, the sentence ‘representations of Venus represent Venus’ is true. In this version, the second-level predicate denotes not the first-level predicate itself, but rather denotes the things the first-level predicate denotes. So it looks as though the explication should be modified to say that

(Two items, a and b, are the same with respect to P_2) *just in case*
(a and b are both denoted by some predicate P_1 , and the members of
the extension of P_1 are denoted by some predicate P_2)

But this creates more problems than it solves. Looking back to the shape examples, the paintings are denoted by ‘rectangular’ but not by ‘shape’; so, in this case, the extension of P_1 includes items not denoted by P_2 , and, thus, on the amended account, the paintings would not be the same with respect to shape. This is, of course, unacceptable. One way of solving this might be to use the amended rule for cases where P_1 and P_2 are the same and use the original rule when they are different. But the *ad hoc* character of this move is not something I’m prepared to live with just yet.

A look at my free movement, up to this point, between predicates and the properties they name might help to shore up this worry. Rectangular is the shape that rectangles have²⁰, so ‘shape’ denotes rectangular, but ‘rectangular’ is a predicate; rectangular things cannot be said to have the predicate ‘rectangular’. This is the difference between a property and the things that have that property. The property is rectangular and the things that have it are rectangles. But this can’t quite

²⁰ Notice that English sometimes obscures this difference: a square *is* square.

be used to unify the account so as to subsume all of the cases under one rule.

Consider the amended rule:

(Two items, a and b, are the same with respect to P_2) *just in case*
(a and b are both denoted by some predicate P_1 , and the property
named by P_1 is denoted by some predicate P_2)

This works for the rectangle case but now excludes the troublesome ‘represents Venus’ case. The three artworks represent Venus, and are thus each denoted by ‘represents Venus’, but the property named by that predicate – something like, representing Venus, or being a representation of Venus – is *not* denoted by the predicate ‘represents Venus’. The sentence ‘Being a representation of Venus represents Venus’ is false.

Given this problem, I see no other way than to handle cases where P_1 and P_2 are the same in a slightly different way from how I handle cases where they’re different. In the cases where they would be the same, there need not be a P_2 at all; so,

(Two items, a and b, are the same with respect to the property named
by P_1) *just in case*
(a and b are both denoted by some predicate P_1)

This second rule, along with the first should include all the sameness relations in any respect. The two paintings of Venus are the same with respect to shape, and they (along with the sculpture) are the same with respect to representing Venus. This version is not quite as *ad hoc* as my prior dividing of the rule, because, with this way, the emphasis is placed on the direct relationship between the objects’ commonality and the respect of sameness. The whole reason for having a second-level predicate in the first place was to explicate the fact that things can be the same

and different *in different respects*, given that certain terms denote them in common. But when the respect in which they are the same is the same property they have in common, then there is no reason to divide the predicate into a first and second level of removal from the objects of comparison.

One final issue I want to address about my latest formulation is the use of ‘properties’ in my explication. A pop psychologist might describe me as having a fear of ontological commitment, so I don’t want to mean anything very robust by the inclusion of properties in my explication. At this point, I’ll just say that I’m using something like a non-naturalized version of what Quine says about ontology, that to be is to be the value of a variable. I count, say, rectangular as a property and real just in virtue of the fact that I can use its name to replace the variable in such expressions as ‘x is a shape’ or ‘some Renaissance paintings are x’ such as to yield true sentences by the substitution. ‘Property named by the predicate’ seems to me just the best name for what I’m trying to name there, despite its extravagant seeming connotations. This extravagance, even if it’s in appearance only, is worth sweeping away if possible, however. I propose to do just this in a way, moreover, that will cooperate well with much of what follows in terms of the application of the explication to cognitive psychology as well as its use in formulating further explications of similarity and analogy. So:

(Two items, a and b, are the same with respect to P_2) *just in case*
(a and b are both denoted by some predicate P_1 , and the property named by P_1 is denoted by some predicate P_2) *just in case*
(a and b are both denoted by some predicate P_1 and P_1 is a P_2 -predicate)

This introduces a systematic notion of a r-predicate, which classifies groups of predicates according to respects of sameness and thus provides a construction that

avoids both the syntactical problems I discuss above regarding predicating predicates and avoids theoretical reference to the vague notion of a property named by a predicate. So, to rehearse examples under this formulation, the Boticelli and the Bronzino are the same with respect to medium, because 'oil' denotes both of them and 'oil' is a medium-predicate. As this illustrates, this move corrects the syntactical problems that arise from ascribing properties to predicates (like 'oil') by formulating a class (r-predicates) designed, unlike other sorts of predicates, to refer specifically to other predicates. As long as we don't think there's anything metaphysically fishy about the notion of, say, a medium-predicate, or a predicate that denotes media, then this move avoids the those kinds of concerns as they pertain to properties. Furthermore, as long as we can also admit that, say, R, is an R-predicate, the move also serves to unify the types of sameness that, previously, had to be treated separately. The Boticelli, the Bronzino and the de Milo are all the same with respect to representing Venus, because they all represent Venus and 'represents Venus' is a represents-Venus-predicate.

Sometimes, explications pay their own way by just being interesting, which I think all of this is, but I also hope more can be made out of it. As I said in first introducing my discussion of sameness and difference, I want to use this explication to ground a further explication of analogy. I will use this further explication to critique the research design and methodology of the long literature within cognitive psychology on cognition of analogy. Before turning to this further task, however, the psychology literature on analogy has developed in tandem with a related literature on similarity itself. I will now discuss how the explication I've proposed and refined

in this chapter so far can be used to elucidate some issues and confusions within that literature.

II. Psychological Similarity: a Possible Application for ‘___ Is the Same as ... with Respect to ----’

In order to show that an explication is worthwhile, it is necessary to indicate its usefulness, a use that we can either recognize as outstanding or as arising with the new possibilities made evident by the explication. A paradigmatic example of such a use is in the clarification or logical streamlining of already ongoing empirical inquiry.

One area of ongoing inquiry where the explicated predicate of sameness might be useful is cognitive psychology. Psychologists for more than the past 30 years have been studying similarity as a theoretical construct that plays a fundamental role in developing models and theories of cognition. There have been scores of studies that attempt to isolate the specific qualities of and generalize over subjective observation of similarity. Nevertheless, psychologists still struggle with questions about whether similarity, as a theoretical construct, is a thing well enough understood (and, indeed, enough of a thing at all) in order to place such studies on a sound empirical footing. In their “Respects for Similarity” (1993), Medin et al. review the cognitive psychology literature on similarity and present research of their own with a mind to showing that similarity *can* signify a worthwhile theoretical construct that shows promise for progress by incorporation into a theory of cognition. The central question they grapple with is one they raise in

response to Nelson Goodman (Goodman 1972): statements of similarity between two things lack informative value without some either explicit or implicit reference to the respect(s) in which the compared items are similar, so similarity *simpliciter*, being too variable to have content, seems to be unsuitable as a theoretical construct. Far from being a challenge to the contentfulness of similarity, the quality of this relation as ultimately variable, especially when we have access to an unbound store of predicates, is one that my own explication of sameness assumes. It is also one that follows from a more basic generalization about similarity, that any two things that are both denoted in common by some predicate are similar. It is common for cognitive psychologists to balk at this generalization, but it is quite obviously true. If two things are both denoted by, say 'F', then they are both F's, and two F's are similar in at least the following way: they are alike in being F's. As long as we are not restricted in which predicates we can choose, it will always be possible to find some predicate that commonly denotes any two things, so any two things will be similar, and if this is the case, then identification of items as "similar" in empirical studies does not pick those things out as different from non-similar things. This radically overdetermined character of similarity is not the exact worry expressed in Medin et al. (1993) (the issue is also taken up similarly in Medin and Goldstone (1995)), but is rather the logical extension of their worry. They worry explicitly about similarity being variable, and so lacking in content without further *addenda*; but similarity is so variable that it relates any pair of items, something that is virtually never recognized in the literature.

So the question that exercises similarity theorists is whether it should be taken as a serious worry about the value of similarity as a theoretical construct that it seems to be infinitely variable. But, unless we reject the claim that any two things that share some property (or unless we restrict the properties we can refer to in an *ad hoc* way), then similarity *is* variable in the way it has been feared. This seems to face us (or cognitive psychologists) with the unwanted consequence of having to exclude similarity as a theoretical construct for use in empirical studies of cognition, although that concept seems unquestionably in use in many cognitive processes. Thus, it's a nice feature of my explication that it both owns up to the radical variability (and, what is more, universal application) of similarity while it also provides other theoretical tools (respects of sameness) that can be used to narrow down this universality to precise formulations that can track the similarities that happen to be salient in particular cases.

The very idea that the variability of actual similarity relations should undermine (or even threaten) similarity as a theoretical construct in psychology reflects a deep and confused equivocation between actual similarity relations and subjective psychological judgments, perceptions, etc. of similarity relations. The history of theorizing similarity in psychology can be read as a story of psychologists making this equivocation and finally getting an inkling of the difference but still lacking a clear notion of actual similarity apart from psychological similarity.

After showing that two notions of similarity are referred to variously and ambiguously in psychological studies of similarity and that ambiguity has led to not a little confusion, I will argue that an objective notion of similarity is needed as a

counterpoint to subjective psychological models theorists want to develop. It is needed, rather generally, because psychological theorists need to have some awareness of the actual phenomena, the psychological correspondent of which they are attempting to model. But, more specifically, a worthwhile account of actual similarity relations would constitute a crucially important tool both in terms of hypothesis formation and experimental design as well as data interpretation. It would also help clarify and dissolve some of the erroneous aspects of previous research, which go unnoticed in the peer-review literature.

My explication of sameness can be used to define an objective notion of the similarity relation that can serve the ends just described. In order to show this, in the next section, I will discuss the difference between actual similarity and psychological similarity, point out where psychologists confuse the two, and argue that psychologists need a to have a notion of actual similarity that they can draw on in studies of psychological similarity. In the third section, I will define similarity and comparative similarity using the predicate ‘___ is the same as with respect to ----’ and show that it meets the relevant criteria for use by psychologists. Finally, in the last section of this chapter, I will discuss a number of cases of actual studies into psychological similarity where the absence of an objective notion is importantly detrimental and the analysis using the notion I develop helps dispense confusion and leads to possible further research questions that do not appear relevant in the absence of that notion.

III. There's Similarity and Similarity; They're Both Alike and Different.

Some attempts have been made by psychologists to answer the variability problem directly, effectively to show, for whatever reason, that similarity is not variable in the ways claimed. None of these are satisfactory, however, because none begin or even attempt to address the more basic (and true) generalization that any two things denoted in common by a predicate (or, in looser language, share a property) are in some way similar. Furthermore, the fact that the evidence that psychologists adduce can only show that subjective psychological notions of similarity are not infinitely variable (and have no bearing at all on actual similarity) suggests a failure to draw a distinction between the two. On infrequent occasions, the literature uses the terms 'psychological similarity', 'similarity in psychological space' and 'subjective similarity', but most often researchers simply use the term 'similarity' as though there were no difference between widespread judgments about similarity²¹ and actual similarity. And, very often, this is not simply a loose way of speaking, as it is an often repeated comparison from William James that has it that both a lamp and a ball are like the moon, but a lamp and a ball are not like each other (James 1892). My objective, in this section, then, is to show a) that similarity is often confused in the psychology literature with psychological similarity, that b) the worry about similarity as a theoretical construct arises as a consequence of this confusion only, and c) that in order to develop cognitive theories using psychological similarity, psychologists need to be able to draw on both notions.

²¹ Or even judgments that aren't so widespread but ones that just happen to be judgments of the researchers themselves.

The general argumentative approach of Medin *et al.* (1993) in support of similarity as a workable theoretical construct is based on empirical studies of similarity, showing that subjects have no trouble picking out things as similar, that there is some regularity in how subjects scale degrees of similarity between things, and that subjects are able to reliably identify some pairs as more similar than others. However, such evidence has no bearing whatever on the questions raised about the variability of similarity. The fact that subjects regard certain things as similar doesn't make them similar any more than the failure of so regarding makes them dissimilar. The only kind of counterexample that would be of use to show that 'similar' does not denote all pairs of items, would be a pair of items that did not have a common property, but there are no such pairs. The fact that many or most of the properties shared by, say, a gnat and a Picasso, are weird and unlikely to occur to anyone gauging the relative similarity of both to a fly and to a Rothko is of no logical significance to whether there are, in fact, such shared properties. If we're talking about the property *smaller than the sun and neither a fly nor a Rothko*, then the Picasso and the gnat are in fact similar to each other (and so much else) but neither is similar to the fly or to the Rothko, at least not similar in whatever respect that weird predicate represents.

Thus, the argument for viability of similarity can be seen to trade in what Frege called "psychologism", for which I suppose we shouldn't fault psychologists too strongly. The problematic step for Frege was to go from general psychological facts about how certain inferences are made to the claim that those facts represent logical laws, or laws about how inferences ought to be made (Frege 1980: viii, ix, 3,

34, 37, 38, 105). Applied to similarity, this becomes the following: there is large-scale subjective agreement about which things are similar to each other, therefore similarity isn't so variable and unconstrained as reflection might suggest. Thus, the considerations regarded as relevant by Medin *et al.* (1993) all have to do with whether similarity is recognizable in example pairs or whether statements of similarity are understood when posed. The most this can show is that some concept of similarity is relied on in cognition, and not that that notion is informative apart from specification of the properties shared or even that such a concept is cohesive at all.

In the paper that Medin *et al.* (1993) and Medin and Goldstone (1995) are reacting against, Goodman compares similarity, relative to some respect²², to motion, which is always relative to some frame of reference (Goodman 1972: 444). Just as whether, how fast, and in what direction an object is moving depends on what frame of reference it is being observed from, the particulars of two items' similarity depend on in what way they are being compared. If this is right, then the operational methodology purported to show that similarity *simpliciter* is a viable theoretical construct is just about as relevant as it would be to obtain intersubjective agreement from subjects observing the sun from their common terrestrial vantage point as evidence to disconfirm the relativity of the sun's motion.

²² Goodman's use of 'respect' in this formulation is importantly different from my own in that it refers simply to the predicate that either does or doesn't denote the compared items. This is not my sense of 'respect', which fits more naturally with common usage. So, Goodman could say that two red squares are similar with respect to red, since both are denoted by 'red', but not that they are similar with respect to color, because neither is denoted by 'color'.

To all of those subjects, it looks like the sun is revolving around the earth, pure and simple.

These considerations point toward the fact that the distinction between actual similarity and perceived similarity (or: “similarity in psychological space”) have been glossed together, confused, or somehow had their roles traded in some of this research. Recall that the worry began about similarity being too variable, drawing on the *conceptual* idea that a shared property constitutes similarity, but then in order to undermine this variability, researchers show and report that subjective judgment about similarity is regular and theorizable. That’s fine, but it does not resolve the conceptual argument from which the discussion begins. However, I do not mean to suggest that actual similarity and psychological similarity are or should be taken to be unrelated; I think the two need to relate in important and specific ways, which I will address below. Before doing that, however, I want to discuss a few points in the history of psychological studies of similarity that highlight the distinction I’m drawing.

One way of looking at the history of theorizing similarity in psychology is as a move from the objective to the subjective, or from the actual to the psychological. A seminal and deeply influential approach to similarity is geometric modeling. Examples of geometric modeling are the multidimensional scaling (MDS) models (Nosofsky 1992; Shephard 1962; Tversky 1977), which input empirical data of various sorts about subjective similarity judgments and output a geometric model of the compared items as points separated in n -dimensional space. The distance between the points is taken to inversely correlate with the similarity between the

items to which the points correspond; so, the closer these points are geometrically, the more similar the things they represent are taken to be. One important and attractive consequence of this approach, then, is that it incorporates a way of evaluating degrees of similarity.

Models of this sort will also not lead to contradicting statements about similarity, and in that sense, they have at least the possibility of reflecting actual similarity relations. But geometric models have been criticized because certain built-in assumptions have been observed to be “empirically violated”. These assumptions are that any two things are less alike than are identicals, that degree of similarity is symmetric, and that degree of similarity preserves the triangle inequality²³ (Tversky, 1977). However, if we look at exactly what these supposed violations are, strange things appear. So, subjects report differing degrees of similarity between “identicals”, but even the posing of the question in that way misuses the term ‘identical’, since actual identicals are not numerically distinct. For example, some “identical twins” may be judged more alike than others, but this doesn’t violate the assumptions of the model, because “identical” twins are not identical; any two members of a pair of identical twins are unlike in countless ways. The symmetry condition is never actually violated either, but the notion that it is comes from Tversky’s famous finding that subjects judge North Korea to be more like China than China is like North Korea. This may tell us something about the racial prejudices of the subjects or provide some clues about their perceptions of international affairs, but it doesn’t tell us anything not so related to the subjects

²³ The triangle inequality holds that the distance between any two points is at most equal to the sum of the distances between each of those two points and any third point.

about China, North Korea or their relation to each other. Thus, such reports have no logical bearing on the model of similarity that assumes symmetry.

The triangle inequality “violation” is much more difficult to comment on without having a worked out account of degree of similarity, but even for this condition, the examples of violation are telling. Such examples always involve two items (a lamp and a ball), which are both similar to a third (the moon) but not at all similar to each other. While it’s not possible to either argue for or against the triangle inequality with regard to *actual* degrees of similarity, without some worked out account of such degrees, this sort of example, on its own, doesn’t constitute much of a violation, since one of its assumptions is false – that a lamp and a ball are in no way similar.

I suggest that we look at the original formulation of geometric models and the subsequent attacks on those models on grounds of “empirical disconfirmation” as a focal point for the issue of actual versus psychological similarity. In some very general ways, the model makes sense as an intuitive and considered stab at actual similarity; considered abstractly, it seems just obvious that things are like each other to equal degrees, for example. So, the original formulation can be seen as one based on the theorist’s own considered view of what actually makes things similar. But, if what the empirical data track is subjective judgments about similarity, then there is no reason to suppose that those data will fit the model, because there’s no reason to suppose that people in general understand actual similarity, regardless of how deep into fundamental cognitive processes similarity judgments (and perceptions) go. There is simply *no reason at all* to make such an assumption. So, the

theory starts out with an abstract model of similarity, then, when the data do not fit the model, the model is revised or rejected. This is appropriate only if the model had been supposed to track subjective instead of actual similarity in the first place. In this instance, it's not clear that these issues were understood as separate, especially since the model designed preserves consistency for judgments about similarity.

Although Gentner and Markman, two of the leading researchers in psychological similarity and (especially) analogy have maintained that “no satisfactory process model for similarity has yet been proposed,” (1995: 111–145) some moves have been made to develop a model that attunes itself to the specifics of how people actually judge similarity, without the reflection of actual similarity relations, which we see much more of in models like the geometric ones. Although no researchers in this field discuss or explicitly recognize the difference between psychological and actual similarity, efforts such as Medin and Goldstone's (1995), which develops a similarity predicate that is relativized in a number of ways, can be seen as attempts to respond with a specifically psychological model to the ways in which psychological tendencies in similarity judgments often fail to reflect actual similarities.

At this point, the question might arise as to what my account of sameness is doing here. If I have made my case about past research and the development of new approaches in response to empirical data, then I've shown that psychologists have gone wrong specifically by worrying about their theoretical constructs and models on account of general truths about actual similarity. If I've made my case, then it

looks like paying attention to actual similarity has gotten psychologists into a fair bit of trouble, so what help could my view be to them?

The fact that cognitive psychologists want their models to track and reflect psychological similarity and not actual similarity, regardless of how wide the former may stray from the latter, does not make an understanding of actual similarity irrelevant for psychologists. If we compare the case to the study of “naïve physics”, for example, the outlines for a use of actual similarity should come into focus. Many of the studies of naïve physics are formulated and inspired precisely by the way in which naïve physics differs from actual physics. Some aspects of mechanics are intuitive and understood without scientific training; some are not. Precisely this divergence focuses important questions in cognitive science and cognitive psychology. Thus, the study of naïve physics draws on actual physics in crucial ways. So, broadly speaking, I will propose an account of actual similarity to function as a counterpoint to subjective models developed to be in accord with statistical regularities in data. I have not yet fully developed my account of similarity as a model ancillary to studies of psychological similarity, but from what I have already said about sameness, it is clear that a model of similarity based on it would preserve symmetry. Given this, what has been cited as a “violation” from the data on the China/North Korea comparison can rather be seen as a divergence in broad intersubjective opinion about relative similarity from actual similarity relations. This is the sort of thing that should spur the asking of significant questions about cognition. Why do so many get this wrong? And how broadly (in the population) does this error spread? Is it in any sense local? Is it a *type* that can be isolated and

summarized? Does it generalize in any way? What effects does it have on other cognitive processes dealing with the compared items? Is the error relative in any way to the format of the question?

In this section, I've discussed worries that have been raised in the psychology literature about similarity as a theoretical construct in general as well as with geometric models of similarity in particular. In both discussions, the issue of actual versus subjective similarity arises, and it appears as though some of this literature confuses and often equates the two. I have distinguished between these two notions of similarity as well as argued in a general way for the need for a good account of actual similarity to draw on in studies of subjective similarity. However, this last point still needs extensive elaboration. I will show, by discussing examples of actual studies, how my account of similarity could help elucidate research both at the stage of hypothesis formation (and, especially, articulation) and data interpretation. But, before doing that, I will develop my account of similarity out of my explication of sameness and show that it meets certain criteria articulated by Medin *et al* (1993).

IV. Similarity from Sameness.

After Geometric models were thought to be unsatisfactory on account of their minimality, symmetry, and triangle inequality, researchers responding to Goodman's challenge that similarity lacks content, began to develop a similarity predicate that takes account of "respects"²⁴. It is important to keep in mind that

²⁴ As well as taking account of so much more; the proposed predicate has the form 'a and b are similar in respect c according to some comparison process d, relative to some standard e, mapped onto judgments by some function f, from perspective g'.

what Goodman and Gentner, Goldstone, Medin and Markman refer to as ‘respects’ are not respects of sameness in my sense. They mean, by ‘respects’ simply the predicates or properties compared items may share. If this is all it means for two things to be the same in some respect, then, the worry that Goodman raises, that similarity doesn’t amount to anything because it simply stands in as an indicator of a shared predicate, has some purchase. However, my explication of sameness in a respect adds a little indeterminacy to this shorthand, which more accurately reflects the openness to interpretation to which statements of similarity are often subject. That is, *contra* Goodman (and the leading psychologists who adopt his maxim), to say that two things are the same in some respect isn’t just to say that some unstated predicate denotes them in common. Respects aren’t shared properties so much as they are ways of categorizing, such as color, shape, size, gender, chemical composition, and so forth. There are at least two considerations recommending this understanding of ‘respects’. First, it fits with ordinary usage of ‘respects’ in a way that Goodman’s version does not. If respects are only shared properties (or, commonly denoting predicates), then students at a women’s college aren’t similar with respect to gender, because no students (at that or any other college) are genders. Following my explication, we can say that students at a women’s college are the same with respect to gender, because they are women, and ‘woman’ is a gender-predicate. What Goodman and those following him call ‘respects’, I have located in my second sameness predicate, ‘___ and are the same in being ----s’. So, the students are the same in being women and the same with respect to gender; thus, my account fits with ordinary usage in ways that the contrasting one does not.

The second reason for preferring my account of “respects” is that it leaves unspecified which predicates the items share (and so correlates with the variability of similarity), yet, at the same time, it *constrains* that variability in important ways, because it indicates which types of predicates denote a pair in common, but does not say how many or which such predicates, always leaving room for interpretive elaboration and investigative discovery. Furthermore, this constraining of variability gives sense to statements about *dissimilarity*, despite the fact that no two things are altogether dissimilar. That is, it allows us to say, for whole sets of predicates, that none of them commonly denote a pair of items.

Medin *et al.* (1993) identify three criteria that they regard as making Goodman’s view of similarity unworkable as a theoretical model in psychology. Although this declaration issues from a failure to mark the difference between actual and psychological similarity, some discussion of these criteria will be nevertheless worthwhile in developing my own view. This is instructive because one of the criteria is purely psychologistic and need not apply to an account of actual similarity, another would have to apply to any account that could interface in relevant ways with psychological investigation, and a third represents a particular and common type of psychologistic error that arises ubiquitously in the literature and follows from a failure to understand actual similarity. These criteria are as follows: any satisfactory model of similarity must a) address the issue of how multiple respects are integrated, b) track the distinction between attributes and relations (and so be able to identify structural similarities), and c) address the

possibility of comparison-dependent and dynamically constructed features (Medin *et al.* 1993: 272).

In this chapter, up to this point, I have been playing loose with the distinction between similarity and sameness, as the explication I developed is of sameness and not similarity, while the latter is the subject of the psychological research I've been discussing. But I think sameness is the more basic of the two relations and can be used to explicate similarity. The basic distinction between these two relations that I will draw is that similarity, unlike sameness in a given respect, varies in degree; so, a first approximation of the view I will develop is that *similarity is the degree to which two things are the same in a given respect*. Two things either are or are not the same in a given respect, but they may also be more or less similar in that respect.

This distinction, between being the same in some respect and being similar, is at the heart of the first criterion claimed by Medin *et al.* to be unsatisfiable by a variable notion of similarity; they say such a notion “does not address the issue of how multiple respects are integrated to determine similarity” (1993: 272).

However, some part of this criterion for similarity is in fact not desirable, and the demand for it as a *desideratum* for an account of similarity as a theoretical construct is partly a consequence of psychologistic inferential moves; that is, we obviously do make distinctions of relative degrees of similarity, but the fact that such distinctions, made in an unrelativized way, can be readily “observed” does not imply that such distinctions track anything objective. Some sense can be given to the notion that two squares are more similar to each other than either is to a rectangle, and both are more like the rectangle than either is to an oval, but that the rectangle is more like

the oval than it is like a circle, and so forth. But when we start adding other respects in which the compared items are the same (respects other than shape), then talk about degrees of relative similarity becomes much less clear, for example, when just one of the squares matches the rectangle in color, is it then more similar “overall” to the rectangle than it is to the square? There seem to be no objective grounds for coming down resolutely on either side of that question, and doing so without some explicit or implied context (that suggests the relevant respect to be, say, color instead of shape) is pointless, regardless of what the “empirical data” from cognitive psychology may show.

So, the first criterion asked for from an account of similarity is partly misguided, but not as misguided as my seem from my remarks just previous; this is because, to repeat again, the term ‘respects’ in that criterion does not refer to the same notion that I have worked out. Medin *et al.*, taking their cue from Goodman, take respects to just refer to predicates that may or may not denote two items in common. This is importantly different from my notion, and notice that, as Medin *et al.* claim, it *does* lack the resources to incorporate multiple properties (what they call ‘respects’), unlike my notion of respects. So, refer back to the square, rectangle, circle and oval. My notion provides a way of ascribing comparative similarity in a given respect:

If one pair of items is denoted in common by more r-predicates than another, then that pair is more similar with respect to r than the other is.

Since “respects” thought of simply as properties shared by two things do not mark out groups of associated predicates in this way, such a conception does in fact fail to

explain how groups of common properties combine to form relative similarities. So, when Medin *et al.* set, as a criterion for an account of similarity, that it should “address the issue of how multiple respects are integrated to determine similarity”, what they really mean is that it should address the issue of how various commonly denoting predicates (representing various respects) are integrated to determine overall similarity.

I have attacked this clarified version of the criterion directly, calling into question the sense of discussing relative overall similarity. But still, some predicates *do* integrate to form obvious senses of relative similarity, like with the two squares compared to a rectangle and unlike the example with an orange square, an orange rectangle and a non-orange square. Clear notions of similarity and comparative similarity can be explicated using my notion of sameness in the following way:

(a is similar to b) *iff* (there is some respect such that a and b are the same in that respect)

iff (there is some r such that a and b are denoted in common by at least one r -predicate)

(a is more similar to b than to c with respect to r) iff
(a and b are denoted in common by more r-predicates than a and c are)

So, suppose we have (a) a red square, (b) an orange triangle, and (c) an orange square, and we have a standard store of shape-predicates ('square', 'rectangular', 'circular', etc.) and a standard store of color-predicates ('red', 'orange', 'blue' etc.). According to the explication, each of the three is similar to each of the others, and

a and c are more similar to each other with respect to shape than either is to b,
because two shape-predicates, 'square' and 'rectangular' denote a and c in common,

but only one denotes a and b and b and c in common. A similar treatment shows that b and c are more similar to each other with respect to color than either is with to a in that respect; b and c are denoted in common by orange, but no standard color-predicate denotes both a and b or a and c.

Insofar as the criterion of integrating a multitude of shared properties can have any grounding in objective facts about similarity, this explication provides a way of integrating them to generate a basis for claims about relative similarity between various pairs of objects. Also, since comparative similarity is always analyzed only with the four-place predicate that relativizes comparative similarity to a specific respect, it seems like the explication does not apply in cases of the ambiguous sort, where it's not clear whether similarity in one respect should trump similarity in another. But this seeming is only partially indicative. This sort of ambiguous case could come up, if we are using an unconstrained store of predicates. Shape and color are pretty standard respects in which we observe, categorize and describe middle-sized objects, and, depending on the context of discussion, there are relatively standard sets of color-predicates and shape-predicates. But what about color-or-shape-predicates? If we were to collect all of those predicates together into the respect "color or shape", then the ambiguity in comparative similarity comes back, except, under the explication, there is a definite answer where it seems as though there shouldn't be one. But color-or-shape isn't a very useful way of categorizing things in general, and it doesn't tend to track salient regularities, so, as a respect, it hasn't become entrenched as a way of grouping similarities, in the way color and shape each have. In this way, the explication of relative similarity, much

like that of sameness in a respect, can be seen to carve out the formal limits of the notion it explicates but at the same time relates substantively at specific points of contact to empirical practice. This is a nice feature of the account for its use as a model of similarity in empirical psychology, since it provides a ground for formulating all sorts of counterfactual questions about the subjective data. Given that, according to the model, there are various answers to the questions of which pairs are more similar, the specific answers given by subjects can be interpreted against the other alternatives of the model to focus other questions about what respects are salient for considering similarity under various conditions and why.

The second criterion set out by Medin *et al.* for a theoretical construction of similarity is more serious; it does concern actual and not just psychological similarity and requires an important emendation of my explication of sameness (and consequently of similarity). Again, they point out that a satisfactory construct would distinguish between attributes and relations, so that similarity based on structure could be incorporated. This is indeed extremely important for any explication that might be offered to aid in current enquiries into the psychology of similarity, not only because much of the similarity literature focuses on structural similarity, but also because theories of similarity cognition are being developed in cooperation with theories of analogy cognition, and virtually all of the dominant current approaches to analogy focus on structural relations. So, given that structural relations are actually distinct grounds for similarity (and not just psychologically distinct)²⁵, unless my account can cover cases of structural similarity, then it will be

²⁵ I will say more about this assumption below.

seriously hindered in its possible contributions to experiment design and data interpretation.

What is meant by “structural similarity” in the stated criterion is similarity based on polyadic predicates. For example, Indiana University is similar to the University of Illinois in that each of the two universities is in the state, a token of whose name appears in the university’s name. ‘Indiana University’ contains a token of a certain state’s name and the university referred to is in that state, and the same with the University of Illinois. There are various structural similarities reflected in these facts. The relations ‘is in the state of’, ‘contains a token of’, and ‘contains a token of ___’s name’ are all isomorphic between the system consisting of {Illinois, the University of Illinois, ‘Illinois’, ‘the University of Illinois’} and the one consisting of {Indiana, Indiana University, ‘Indiana’, ‘Indiana University’}.

One way that my account of sameness (and similarity) could accommodate these relational similarities is by simply identifying the relations as complex monadic predicates; there is no purely logical reason why this couldn’t be done. So, for example, the two universities are the same (and at least minimally similar) by virtue of the fact that the predicate ‘is in the state a token of whose name appears in its name’, what we might call a relative-location-predicate, since it specifies a thing’s location in relation to some other thing. These cooked-up monadic predicates capture facts that statements of relational similarity report, but they do so in a way that significantly diminishes their informative value and formally undermines many of the inferences in which we might want to use these facts. For example, given that Indiana University is in the state of Indiana, it follows deductively that there is some

state and Indiana University is in it. But, if I have to schematize this relation using the proposed predicate ‘is in the state a token of whose name appears in its name’, then this inference cannot be drawn, because the second item of the relation, the state, Indiana, is built into the predicate and not signified logically as a variable.

One reason, though, that moves of this sort might seem like a good first attempt is due to the way the criterion is framed (as is discussion of structural similarity generally in the psychology literature); that is, structural similarity is ascribed to individual items, when, in fact it obtains between groups of items, or systems. So, if you are actually looking for what’s similar between the University of Illinois and Indiana University in terms of location relative to state, then the monadic predicate is correct, since it ascribes a common property to the two things that are supposed to be similar. But, the interesting similarity is not between the two individuals; it’s between two groups of items. My explication of sameness can be emended to accommodate this latter sort of similarity in the following way:

$\langle a, b \rangle$ are the same as $\langle c, d \rangle$ with respect to relative- r iff
 (there is at least one relative- r -predicate that denotes both $\langle a, b \rangle$ and $\langle c, d \rangle$)

Using this emendation, $\langle UI, \text{Illinois} \rangle$ and $\langle IU, \text{Indiana} \rangle$ are the same with respect to relative-location, because the relation ‘is in’ denotes both ordered pairs and is a relative-location-predicate, a predicate that ascribes location relatively. The other similarities of the two university/state/name systems work similarly. The pairs $\langle \text{‘University of Illinois’}, \text{Illinois} \rangle$ and $\langle \text{‘Indiana University’}, \text{Indiana} \rangle$ are both denoted by the relation ‘contains a token of ___’s name’, and so forth. Also, it is obviously

possible to treat more-than-two-place predicates similarly, so that all sorts of individual structural relations can be ascribed.

The final stated criterion for a theoretical construct of similarity is that it must “address the possibility of comparison-dependent and dynamically constructed features.” One of the ways that psychologists have devised for coping with the inconsistencies that arise when they start looking at subjective judgments of similarity in test groups is by incorporating contextual elements into their models, such as the order in which two items are compared with a third. Below, I will discuss a well-known experiment showing that changing such ordering correlates directly with changes in judgment about what is similar to what. To accommodate and attempt to explain the inconsistencies that arise as a consequence of this phenomenon, psychologists have suggested that properties themselves may sometimes be context-dependent and “constructed” on the fly during acts of cognition. Some properties and relations are, in fact, context-dependent, such as spatial and temporal properties and relations and whatever property is referred to by the predicate ‘referred to in the present context’, but these aren’t the types of properties that psychologists are worried about. Thus, this concern is a prime example of how an understanding of actual similarity, as a factual relation, could help clarify and motivate past and new research. That is, psychologistic similarity has led to absurd conclusions about the context-dependence of properties, instead of the other way around; the fact that psychologistic similarity plays loose with and constantly adjusts concepts in ways

that lead to immediate inconsistencies should be a jumping off point for inquiry into the specifics of those adjustments.

My discussion of this last criterion and “context-dependent properties” has been fairly abstract, but I will return to it in a more concrete way in the next section. In the next section, I will discuss a number of actual studies, comment on what they were taken as showing and show how these studies could benefit from a confrontation with my account of similarity.

V. Empirical Points of Contact

A. “Ambiguous” Cases

One of the studies that has been taken to suggest the notion that the having of properties is context-dependent uses the following, Figure 1 (Medin *et al* 1993):



Subjects are asked either to compare A and B or B and C. B is described by designers of the experiment as “ambiguous” as between having three and having four prongs. If it has three, then it’s similar to A with respect to prong number, and if it has four, then it’s similar to C in that respect. Researchers found that when comparing B to A, subjects were far more likely to attribute three prongs to it than not, and when comparing B to C subjects were far more likely to attribute four prongs to it than not. The projective response is that the having of properties may be dependent upon

historical facts about comparisons. So, certain properties of B are relative to B's having been compared to A or having been compared to C, for example.

There's a fair bit of equivocation going on here, since the proposed fix, to make properties relative to comparison, is undermined once we let the two subject groups talk to each other. A/B comparers will surely disagree with C/B comparers about how many prongs B has. If the judgments are to be made consistent, then two notions of 'prong' must be at work. Since my account of similarity is based on predicates and takes predicates to be functions mapping objects onto truth-values, it provides a focus for this line of research that does not trade in equivocation or in subjective, fleeting properties. According to this account, which takes predicates to have definite extensions, two notions of 'prong' must be at work in the different groups of subjects, the group that compared B to A used a notion that excludes the rightmost element of B, and the group that compared B to C used a notion that included it. Since my account is based on predicates, not on words or concepts, the difference between the two groups can be accounted for by attributing the use of different predicates for "prong" by each of the two groups. This also provides focus for further research questions about concepts and applications of them, since it seems to suggest that a concept, like "prong" is intersubjectively underdetermined in terms of correct applications.

B. Symmetry

As mentioned above, the geometric models of psychological similarity first came under fire, in part, as a consequence of empirical data showing that judgments

of similarity are often asymmetric. I mentioned the China/North Korea case, but there are many other pairs that have been shown to regularly elicit similarly asymmetric judgments of similarity. Such pairs are pencils/crayons, England/United States, Skateboards/Bicycles, and Albert Einstein/Benjamin Franklin (Medin and Goldstone 1995). The general theorizing of these data has let researchers to conclude that similarity between pairs is direction-sensitive.

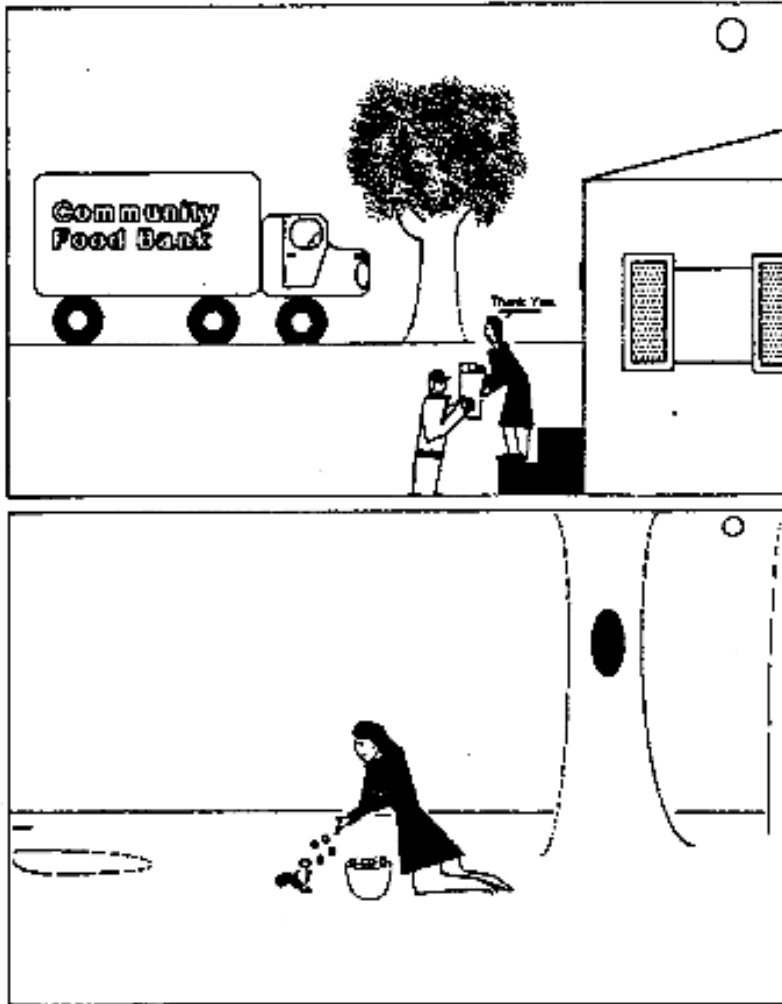
It may be the case indeed that subjective judgments of similarity between pairs are direction-sensitive; nevertheless, if we consider this in comparison to actual similarity, which is not direction-sensitive and is symmetric, it becomes a focal point and experimental control, because it represents widespread error and confusion on the part of research subjects.

One thing this sort of error may suggest is a lack of equality in psychological salience between the two pairs. In all of the pairs mentioned, at least for subjects in the United States, which is where the studies were conducted, it seems plausible that one of the pair's members is simply more significant psychologically, either because more is known about it, it has more prestige, appears more often in everyday life, or some other reason. Another example of a direction-sensitive pair, butchers/surgeons, would bear out this hypothesis. The butchers/surgeons pair is not offered in the literature so much as an example of asymmetry of degree of psychological similarity but of the direction-sensitivity of the quality of similarity (Medin *et al.* 1993). Thus, comparing butchers to surgeons compliments butchers, but comparing surgeons to butchers disparages surgeons. This example would also conform to my salience hypothesis.

So with the help of a comparison to actual similarity, the subjective similarity data can be mined for interesting and promising research initiatives. Which comparison types give rise to asymmetric similarity judgments (in degree or quality) – or, more correctly – *misjudgments*? What roles do these types play in cognition more broadly, especially, as compared, say, to types that do not give rise to such asymmetries? What fitness advantages, if any, accrue to misjudgments of similarity of this sort? What are the likely consequences of undermining the tendencies for this type of misjudgment through educational initiatives?

C. Similarity and Alignment

The golden key of studies in the cognition of similarity seems really to be an understanding of the cognition of analogies, and more and more studies of similarity are focusing on what researchers call ‘structural similarities’, which are taken to represent analogical rather than feature-based similarities. The basic hypothesis is that in judging similarity, global, and thus, structural similarities are more important than local (sometimes called “attributional”) similarities. Figure 2 is used in studies to support this hypothesis (Markman and Gentner 1990):



The basic finding is that when asked about correlations between the pictures and what “goes with” what, subjects tend to correlate the woman in the top picture with the squirrel in the bottom picture, despite the fact that the women in the two pictures are similar in appearance. This is taken as indicating a focus on structural similarities over attributional similarities. Thus, the giving/receiving relation is the one that subjects find salient in identifying what “goes with” what, instead of the attribute of being a woman, having dark hair, wearing a dark dress with long sleeves and a short hemline, etc.

One problem with this study is the wording of asking what “goes with” what, since this connotes a structural relation (as compared with an attributional similarity) in the first place, a connotation which surely biases the data. But beyond such operational matters, there are more fundamental theoretical issues involving actual similarity that must be worked out before an approach to such questions could be legitimately made. The general shortcoming of this, as with virtually every word of discussion on the matter of relational versus local (or attributional) similarities in the psychology literature is that it trades on indeterminacy in the compared items. No attempt is given to make sense of the distinction between ‘global’ and ‘local’ matches, but it is virtually certain that ‘global’ and ‘local’ do not track the difference between structural and attributional relations.

One sense that could be reasonably given to the global/local distinction is that global predicates denote the whole item and local predicates denote things contained in the item. So, a global predicate true of both drawings in Figure 2 would be ‘depicts a woman’; a local predicate true of the woman in the first drawing but not the second is ‘standing’. In this context, ‘standing’ is a local predicate, because it’s true of the woman in the drawing, but not of the drawing itself: the drawing is not standing.

This might not be the only reasonable sense of global versus local, but there seems to be no sense in which structural properties are more global, in the sense that they are true of the “whole thing”, than simple (one-place) properties are. In fact, structural properties, since they are indicated by multi-place predicates, are, in some sense, *never* global, because they are never just about the item being

compared; they are about elements within that item or that item's relation to things outside itself.

But suppose this is just an unfortunate conflation of 'global' with 'structural' and 'local' with 'attributional'. Giving up these conflations doesn't clear up the nascent confusions; rather, these conflations indicate a heavy reliance upon the indeterminacy of the compared items. That is, in comparing the two drawings, we can take them as wholes, considering only those predicates that might be true of them as a whole, like 'depicts a woman', 'drawn in black ink', and so forth. Or we might consider them as systems with three elements, a, b, and c, such that one of the three gives a second to the third. Or, we might consider them as systems with any other number of elements.

One of the interesting features of psychological similarity surely has to do with the way this indeterminacy becomes focused into specific structures in the act of cognition. The pictures themselves can't tell you what to see them as; simply, they are just so many things. Some of them might be so obvious as to get noticed "automatically", as it were; but others may not. Such questions seem to be at the core of studies like the one I'm discussing now, where the fact that there are different possible ways of reading the two items is more obvious than in many contexts. However, due to the lack of recognition of the indeterminacy of these items being compared and lack of awareness of the deep difference between the picture and various structures that it might instantiate, these questions cannot be posed in a rigorous way.

My account of actual similarity provides the resources for posing such questions rigorously, because, especially in the definition of similarity between pairs of items, it makes clear exactly what is being compared. So, under my analysis, the statements of actual similarity between, say, the two pictures “attributionally” and “structurally” would look quite different, because they would not compare the same things. In a loose sense, we could say that both compare and report similarity between the two pictures, but, in a strict sense, it is not really the pictures that are being compared. In the “attributional” case, it’s the women who are being compared, or perhaps it is the relation (of, say, depicting) between the pictures and the women in them. In the structural case, it is the donor, the food and the recipient in each picture that are being compared, or perhaps the relation between each of the pictures and the donor, food, and recipient. So, in statements of actual similarity, the indeterminacy that accompanies comparison of items in the world is specified, and we see that it is not so much (or, not only) a *type* of similarity that is more often chosen between the two pictures, but is also a way of looking at the picture in the first place, as a thing containing or depicting a set of items and their interrelations, instead of containing a set of items having various attributes. My account of actual similarity both provides the framework for this interpretation of the data as well as provides a theoretical basis for designing future experiments so that these questions can be rigorously investigated.

The confusions arising from failure to recognize this indeterminacy are pervasive. And properly addressing them involves more than just pointing out that comparing items often involves not really the items of stated comparison but things

in some sense “within” those items. This issue of plurality is only half of the story; the other half has to do with predicates included in structures. Here is another example of the same error from Medin and Goldstone that will help me get at this other half of the issue:

Let drawing *E* be a triangle above a circle, Drawing *F* be a circle above a triangle and Drawing *G* be a star above a square. Both *E* and *F* have a triangle, a circle, and the relation *above*. *E* and *F* are not, however, identical and we may try to capture this difference by treating the arguments (circle, triangle) as bound to the relation *above*. [...] On the other hand, *E* and *F* also resemble *G* in that all three contain the abstract relation of aboveness. To capture this similarity, we would want to treat the *above* relation as unbound means that the order of the arguments within the relations is not relevant. One could describe the situation in terms of attributes, bound relations, *and* unbound relations but to do so would be to abandon the idea that features should be independent and non-decomposable (or at least not a function of features in the same descriptive language). We see no alternative to directly addressing structure, even if it entails giving up the idea of independent features (1995: 98).

The three drawings, and anything else we might care to evaluate, instantiate a countless number of structures. Just as the study of property similarity focuses largely on the framing of properties and property sets, the same will hold for structural similarity with regard to structures. One structure shared by *E* and *F* contains the predicates ‘triangle’ and ‘circle’ only; another contains ‘above’ only and is also shared by *G*. A third, contains ‘circle’, ‘triangle’, and ‘above’ and is shared between none of the pairs. The problem again is that drawings cannot tell you what they contain; they contain so many things. So, to say, for example, that *E* and *F* contain circles and triangles (which is effectively what it means to treat them as “arguments”, which are supposed to be free variables) is to obviate the possibility of making similarity judgments based on structures which do not contain the

predicates 'circle' and 'triangle'. Once we realize that there are many, many structures instantiated by any object, and in this case, at least a few interesting ones, then confusions of the sort in the quoted passage pass away.

However, it might be maintained by the psychologists or by some metaphysical realists that, although there are myriad structures in the way that I'm suggesting, some are fundamental and some are not. That, for example, it is a fundamental fact that E and F contain one triangle and one circle each, so my account, which draws on a shared structure between E and F which contains neither 'triangle' nor 'circle', fails to fully comprehend the true structure of E and F. The best sense that I can make of such a reaction is as one that is psychologically motivated: it is difficult to cognize such drawings as E and F without noticing that they each contain a triangle and a circle. This may be, but that does not give primacy to facts about E and F *vis-à-vis* triangles and circles in terms of *truth* over other sorts of facts about E and F that do not recognize those items as triangles and circles. So, for example, it is no *less true* that E and F both contain two shapes, and that they have ink marks on them, that each contains three angles, all adding up to 360 degrees, and so forth. Given that the difference in interpretations of the drawings cannot be cast in terms of truth, some other way would have to be given, although I don't know of any promising attempts.

So, bringing my account of similarity to bear on studies into structural alignment has shown that a great deal of loose talk has gotten worked into the formulation of these studies and the interpretation of their data. My account provides an objective model of similarity not only between individual items but also

between systems and thus theorizes the difference between various structures that a given item (and its components) may instantiate. Getting clear on these issues looks to provide a promising guide for posing interesting questions that follow from the model, such as questions about which sorts of structures tend to become cognitively apparent in perception, whether the structures of focus vary under certain conditions, and where the upper limits of complexity tend to lie. It's obvious that once a structure becomes too complex then it will be difficult to perceive and use in cognition (in this case, comparatively), if possible at all. That's what a good part of IQ tests attempt to track. This, of course, conflicts with the general claim of the study involving the two drawings; once there are too many elements, and the relations between them are too complex, then more easily perceivable features are likely to be chosen to "go with" each other. The fact that my account helps get clear on the relevant issues – number of components and which predicates are included – for understanding shared structures, makes it a useful tool in focusing research on questions such as the one I've just suggested.

D. Missing Body Parts

A final aspect of psychological similarity and its study that I will discuss has also to do with alignment. The general idea is that what counts as a match or a mismatch between items depends on facts about comparisons. Examples such as the following are offered in support of this notion: Suppose we are comparing drawings of people. Drawing A is missing a hand, Drawing B is not missing any body parts, Drawing C is missing a hand and a foot and Drawing D is missing a foot. It is alleged

by Wisniewski and Medin (1990), that, although A is more similar to C than to B, because A and C are both missing a hand, A and C are dissimilar in that C is missing a foot, while A is not. All of these comparisons are based on what they call “frames”, by which they mean alignment of a part with another of its type. So, the hand, or missing hand is aligned with a hand or a missing hand, and so forth. Once D is brought into the picture, things change, because D is taken to be more like A than C is, because D is missing exactly one appendage, like A, although that appendage is a foot in the case of D, while a hand in A’s case. The alleged conclusion from this is that two mismatching “frame-based” comparisons (missing hand to hand, and foot to missing foot) are converted into a “higher-level” matching comparison.

The distinction between “frame-based” and “higher-level” here, is psychologistic, and illicitly so, since it is supposed to be a theoretical distinction, one in terms of which data are interpreted. Nothing more is meant with ‘frame-based’ than that the correlation is obvious, but obviousness, because it is subjective, does not make a useful distinction for investigation. It may be more readily apparent that the missing thing in A is a hand and, in D, a foot, than it is that both A and D are missing appendages, but the latter is no less true and is not, for its lack of obviousness, a “higher-level” fact. Nevertheless, just as with all of the other examples I have discussed, interesting things are going on in this example that can be brought out using my account of similarity.

Applying my account, the difference between the A/C comparison and the A/D comparison is in terms of respects. A is similar to C with respect to hand-number, for example, and it is different from D with respect to hand-number. Yet, A

and D are the same with respect to appendage-number and A is dissimilar from both B and C in that respect. In this way, the issue of respects²⁶ comes into focus as a topic of research. While the proposed conclusion from this example was that matches and mismatches are comparison-specific, the theoretical distinction used to clarify the move from one to another type of matching as a consequence of comparison was not worthwhile. Respects of similarity provide such a way in that they provide a way of distinguishing what is being compared in terms of what type of thing it is evaluated as. In this way, and unlike the “frames”/“higher-level” distinction, it does not conflate the psychological (the readiness-to-mind of ‘hand’ over ‘appendage’) with the conceptual. Still, the conceptual application provides controlled space for empirical research, the move from the respect of hand-number to appendage-number represents a shift in perception, and, given the data to suggest that A and D are “more similar” than A and C are, it seems to represent an important shift in perception. Respects provide a theoretically grounded way of articulating the differences across such shifts.

VI. Conclusion

In this chapter, I have pursued three objectives. First, I have developed an explication of the notions of *sameness*, and *difference*. I have discussed possible problems with these explications, have refined them where needed, and have clarified important characteristics and consequences of each. Next, I have argued that much of the psychological literature conflates subjective notions of similarity,

²⁶ In my sense, not that of the psychological literature, as discussed above.

especially notions that develop out of statistical regularities in their subjective data sets, with actual similarity. I have argued that some of the theoretical concerns about similarity on the part of psychologists are, thus, misplaced, because they have to do with actual similarity while the data that seem to contradict them are subjective. My second aim in this has been to reverse this relation, instead of *worrying* about the fact that subjective data often contradict basic conceptual facts about similarity, it would be useful to exploit these contradictions for the purposes of inquiry. To this end, I have developed an explication of similarity out of my explication of sameness to serve as a comparison to models of subjective similarity that are developed from statistical regularities in subjective data. Finally, I have discussed several studies and arguments about those studies from the psychology literature on similarity, where I have used my own explication of similarity and its general predicate theoretic framework to point out problems with data interpretation and to suggest possible solutions and other possible paths of inquiry that are made possible and recommended by this modified perspective.

CHAPTER 4

ANALOGY AND PSYCHOLOGY

In the previous chapter, I offered an explication of the sameness relation to develop an explication of similarity such as could be useful in studies of similarity in cognitive psychology. I also argued there that such studies suffer from the conflation of and equivocation between a psychological notion of similarity and actual similarity. In this chapter, I will provide a similar treatment for analogy, as a special type of similarity. This also follows the main line of investigative inquiry in empirical psychology, as much of the reason for focusing on similarity in the first place is in an effort to understand the cognitive role of analogy, taken variously to be a more central, more basic, or more cognitively substantial variety of similarity. Some research has even been offered to suggest that judgments of simple similarity tend to focus on analogical, or structural, features over “surface” features. So, it may be the case that analogy represents the more central and important cognitive process; and it may turn out that its mechanism underlies the more pedestrian process of cognizing simple similarity. Given these considerations, whatever ground I might have gained in critiquing and patching the methodological assumptions and apparatus of cognitive psychology’s inquiry into similarity in the previous chapter, the stability of that ground can be relied upon only to the extent that it holds solid as a foundation for a further foray into the domain of analogy itself. Given these considerations, in this chapter, I will use my explication of similarity to develop an explication of analogy, or analogical similarity; then, I will outline the research on analogy in cognitive psychology and critique this research by the lights of my own

explication of analogy, showing where this explication can assist in resolving errors, clarifying confused research methodologies, and suggesting promising new hypotheses and avenues for inquiry.

I. An Explication of Analogical Similarity

The term ‘analogy’ has a fairly broad application in ordinary speech. It can be used to specify a similarity in ratio, or proportion, a similarity between internal structures, or a similarity between two things sharing more “surface”-type features. For my part (and, similarly, for the purposes of cognitive psychology), ‘analogy’ will refer specifically to relations of similarity with respect to structure.

A pair of consequences of this focus is that 1) some things that are called “analogy” aren’t, by my lights, analogies in any interesting sense, and 2) some things simply called “similar” are analogies. For an example of the first consequence, a film featuring a troubled prince might be said to be analogous to *Hamlet*, while no deep, interesting structural features are shared between the two works²⁷, such as similarities in the cause of the trouble, and the relations between the other characters and the prince and his troubles. Such similarities, based on shared features, such as can be identified by denotation-in-common by a one-place predicate, don’t signify analogies in any interesting sense, regardless of whether such cases might be covered by an ordinary language understanding of ‘analogy’.

Furthermore, one thing that motivates this distinction is the fact that it can be

²⁷ I say ‘deep, interesting’ at this point as a bit of intended imprecision. The problem here, as is crucially overlooked without exception, is that any two systems share structural features; they may not, however, share such features in a way that is obvious or even accessible without a great deal of effort. I’ll say more about this below.

specified schematically, that is, in terms of the difference between the common denotation of a pair of objects by a single one-place predicate and an isomorphic mapping of a set of objects onto another set in such a way as to preserve relations between the members of each set *vis-à-vis* some set of polyadic predicates. A similar distinction has been relied upon in empirical cognitive psychology; however, in that domain it has been relied upon in untenable ways, which I will address below.

As an example of the second consequence mentioned in the previous paragraph, we can look again to a perturbed prince and the company he keeps. One might reasonably say, regarding *Hamlet*, that Laertes is like or is similar to Fortinbras. Though the indirect statement uses the term ‘like’ or ‘similar’, it can still be read as an analogy in my sense. It makes most sense to read it that way, given what we know about the two. Laertes and Fortinbras aren’t similar in many interesting superficial or surface aspects. Laertes is Danish, a student in France, has a sister, is a son of a councilor; Fortinbras is a Norwegian, a prince and general, and the son of a king. However, within the structure of the play, Laertes and Fortinbras are similar in their relations to each other and to others in the play. Laertes and Fortinbras both want to avenge their fathers, and, in some sense want to strike an aspect of Danish power in order to do so. And, with respect to Hamlet himself, they stand in unique comparison: they are both active, by comparison to Hamlet, but the one acts with pragmatic judgment, while the other acts without such judgment. All of these similarities rely on relations to which each of the two, Laertes and Fortinbras, stand with respect to other elements of the play. Since this is the case, it is possible to specify a pair of structures, both instantiated by the play, one involving

Laertes and the other involving Fortinbras such that the one structure would map onto the other. For example, one pair of relations, 'more active than' and 'more pragmatic than', have specific truth values when applied to the set {hamlet, fortinbras}, and these relations can be mapped onto the pair of relations, 'more active than' and 'less pragmatic than', as they hold of the set {hamlet, laertes}. Every appearance (and only those appearances) of fortinbras in the extension of a relation in the one structure is matched by an appearance of laertes in the other structure.

So, in effect, and following my methodology of explication, statements purporting similarity may well report significantly analogical relations, and statements purporting analogies may be used to report only surface similarities.

My arguments in the previous three paragraphs are underwritten by a distinction that is widely overlooked in ordinary speech and is usually ignored, unnoticed or underplayed by researchers studying analogy. That is, these arguments are underwritten by the notion that similarities hold between individual objects and analogies hold between sets of objects, or systems. One way of filling this out is to consider whether it would be possible to construe the analogical relation between Laertes and Fortinbras discussed above in terms of one-place predicates instead of relations (since relations make reference to more than one object. It turns out that it is not possible to represent the same informational content in the Laertes/Fortinbras analogy without reference to other objects (and thus without the use of polyadic predicates). We could, for example, say that Fortinbras and Laertes are both more active than Hamlet (both are denoted by the one-place predicate 'more-active-than-Hamlet'), but this loses the relative difference in their

pragmatism that makes the analogy helpful for interpreting the play. The predicate 'less-pragmatic-than-Hamlet' denotes Laertes and 'more-pragmatic-than-Hamlet' denotes Fortinbras, but these predicates are different and so don't form a basis for similarity. A one-place predicate that could function for this latter purpose (forming a basis for similarity) would be 'either-more-or-less-pragmatic-than-Hamlet'; this predicate describes a way in which Laertes and Fortinbras are similar, *vis-à-vis* their pragmatism and Hamlet's, but it loses most of the content of the original analogy – most, if not all of the characters (other than Hamlet himself) are denoted by it. To incorporate the structural similarity at which the analogy aims, it is necessary to use polyadic predicates, and thus, to employ reference to a multitude of objects; that is, not just the pair {fortinbras, laertes} but rather the sets {fortinbras, hamlet}, {laertes, hamlet} must be referred to. Since this is the case, talk of two things being analogous to each other is virtually always subject to some more precise formulation to bring out the content of the analogy. Either the two things compared are items in some larger group, as in this case, or the two things themselves represent groups of objects, such that these objects and the relations between them need to be identified in order to bring out the content of the comparison. This need is not explicitly understood in research on analogies, and it is therefore not reflected in the methodology and interpretation of such research. This failure leads to considerable confusion in these pursuits.

In my discussion, up to this point, I have been speaking as though there were a *single, specific content* that an analogy or a similarity comparison instantiates. This is not strictly the case, but, since statements of analogy and similarity are non-

specific such that they underdetermine relations that may be reported by them, it is difficult to avoid this problem in ordinary speech. Thus, when I talk about *the* analogy between Laertes and Fortinbras, I don't mean (and I'm not committed to the notion that) there is only one such analogy or that every utterance of the sentence 'Fortinbras is similar to Laertes' picks out the structural similarity that I've been discussing. The similarity meant may well not even be a structural one (thus, not an analogical one), rather, depending on any number of factors, the similarity meant may be something more attributional, such as the fact that (after Act III scene iv) both characters have dead fathers. Furthermore, just as such a statement cannot on its own tell you whether an analogical (structural) or a surface similarity is meant, it also cannot on its own tell you which of any number of analogical (or surface) similarities is meant. The underdetermination is not, then, just between an analogical and a surface similarity but also between any number of specific analogical or surface similarities. I have focused on the analogical relation I have, summarily the notion that both Laertes and Fortinbras act as foils to Hamlet, with respect to his willingness to act and his pragmatism about action, because this is a central part of one common reading of the play, and something that someone who has something interesting to say about the play might mean by the remark that Fortinbras is similar to Laertes.

Just as the fact that analogies hold between or at least entail the reference to sets of objects (instead of pairings of single objects) gets obscured in the analysis of analogies in most empirical studies of analogy in cognition, this latter point, that reports and statements of similarities and analogies underdetermine which of any

number of relations might be meant, also has remained obscure up to this point. So, it is usual in such studies to hear talk of *the* analogy between a and b, when any number of analogies hold between them. We may tend to focus on particular analogical relations in practice, such as I have with the example of Fortinbras and Laertes, because of some aspect of our purposes, such as that it represents one aspect of a common reading of the play, but it is nevertheless the case that many other analogies hold in this, as well as any other example. So, another such structural relation between Fortinbras and Laertes can be formulated in the following way. If we take the set {shakespeare, fortinbras, old norway} and the set {shakespeare, laertes, ophelia}, the extension of the relations 'wrote a play involving', and 'was the father of' from the first set map isomorphically onto the extensions of 'wrote a play involving' and 'was the sister of' from the second set. Given these sets of relations (and, thus, this pair of structures), Fortinbras is structurally similar to Laertes; everywhere Fortinbras shows up in the first structure, Laertes shows up in the other. But, as it stands, there's nothing too interesting about these structures on their own or their relations to each other. They amount to just saying that Shakespeare wrote a play involving Old Norway and Fortinbras, and that the one is the father of the other, just as he wrote a play involving Ophelia and Laertes, and the one is the sister of the other. By iteration of similar shared structures that could be generated at will as freely as the facts allow (and there are *always* facts, though they might be weird, random, or irrelevant), this example shows that the mere identification of a structural relation can never be enough to explain an analogy (or anything else), since such relations are always

available to those with a little ingenuity or, sometimes, just an eye for circumstantially connected sets of banal truths.

Let me take stock. As the foregoing remarks suggest, by offering an explication of analogy, I do not mean to suggest that such an explication can, on its own, explain what is meant either by a statement of analogy or of similarity or by a particular utterance or inscription of such a statement, because such statements underdetermine which similarities are being pointed out, and, often, they even underdetermine reference to what is being compared. Rather, the explication should provide a schematic device that can clarify the relation, whatever it may be. In this sense, the explication provides a framework within which relevant clarificatory questions can be raised and thus provides a theoretically grounded starting point for investigation into utterances reporting analogy and other types of indications of the cognizing of analogies.

Analogy, as a type of sameness between groups of objects, can be explicated using the explication of sameness in a respect that I developed in Chapter 3, because analogical similarity can be analyzed as sameness with respect to structure. Or, filling this out:

1. (a is analogous to b) *just in case*
2. (a and b are the same with respect to structure) *just in case*
3. (there is some structure S, such that a and b both have S) *just in case*
4. (a and b are denoted in common by some structure-predicate) *just in case*
5. (for some set of polyadic predicates denoting n -tuples of the elements of a, there is a set of polyadic predicates denoting n -tuples of elements of b, such that each element of a can be mapped 1-to-1 onto an element of b) Or: a and b are structurally isomorphic.

This explication introduces theoretical notions of “having a structure” and of a structure-predicate. The structure of an object (the structure it “has”) can be summarized by identifying a set of polyadic predicates and a specification of how the elements of that object appear in the extensions of those predicates. So, to return to the example I’ve discussed up to this point, suppose we’re talking about what might be called the Hamlet-Fortinbras Complex, a thing composed of the members of the set {hamlet, fortinbras}. One structure this instantiates can be summarized by specifying the relations ‘more active than’ and ‘less pragmatic than’ and the fact that (ranging over the Hamlet-Fortinbras Complex) the extension of the first relation is {⟨fortinbras, hamlet⟩} and the extension of the second is {⟨hamlet, fortinbras⟩}. However, I am careful to call this a “summarizing” of the structure and not the structure itself, which must be taken to be more abstract (and thus something that is not defined by the specific relations stated). This level of abstraction is needed if items can be said to “share a structure”. That is, if the structure to be shared is itself the pair of relations, ‘more active than’ and ‘less pragmatic than’, and the way in which the members of the set fit into the extensions of the relations, then this structure is not shared with the Hamlet-Laertes Complex, because Hamlet is not less pragmatic than Laertes. This would undermine an isomorphic mapping, as follows:

elements: {hamlet, fortinbras}	elements: {hamlet, laertes}
more active than: {⟨fortinbras, hamlet⟩}	more active than: {⟨laertes, hamlet⟩}
less pragmatic than: {⟨hamlet, fortinbras⟩}	less pragmatic than: {⟨laertes, hamlet⟩}

This way of fleshing out the notion of a “shared structure” prevents an isomorphic mapping. Taking just the first relation, hamlet maps onto hamlet and fortinbras maps onto laertes, but including the second relation reverses the mapping and undermines structural isomorphism. So, at least in this case, the shared structure cannot be the relations and their extensions in a literal sense. However, if we consider these relations abstractly, for example, using dyadic predicate schemata, R and T, then the two “complexes” can be said to share a structure in the following way:

elements: {hamlet, fortinbras}	elements: {hamlet, laertes}
R: {⟨fortinbras, hamlet⟩}	R: {⟨laertes, hamlet⟩}
T: {⟨hamlet, fortinbras⟩}	T: {⟨hamlet, laertes⟩}

In the first column, R, can be replaced by ‘more active than’, and T can be replaced by ‘less pragmatic than’; in the second column R can be replaced by the same (‘more active than’) while T can be replaced by ‘more pragmatic than’. These replacements generate true sentences and preserve isomorphism between the complexes or systems. So, what is meant by “sharing a structure” is that an isomorphic mapping of elements interpreted under some set of polyadic predicates onto elements interpreted under some (perhaps other) set of polyadic predicates. Since the predicates need not be the same, the structure itself is abstract and can be identified using substitutional predicate schemata and their extensions. Working this into my notion of sameness in a given respect involves admitting that any pair of systems

that are isomorphic in this way “share a structure” and, thus, are denoted by some structure predicate.

There is one modification that needs to be made, in order for the explication to apply more broadly, however. According to the letter, the account of the analogical relation given up to this point doesn’t apply directly to Fortinbras and Laertes; notice that when I started to apply the analysis, I switched from talking about Fortinbras and Laertes themselves to talk of the “Hamlet-Fortinbras Complex” and its analogical counterpart. This was necessary, because, in the terms of the explication, the items of analogical comparison, the analogs, are sets of items, or systems with various elements, not singular objects. But many analogies, as stated, are not between systems as wholes, such as the “Hamlet-Fortinbras Complex”, but between individuals, such as Fortinbras, which (or, in this case, who) represent elements of such systems. One way of answering this issue would be to say that analogies are *really* between systems, not individuals, and that statements of analogical relations between individuals are always elliptical for analogical relations between systems of which those individuals are a part. However, although analogies always involve systems, there seems to be reasonable sense to be made of the focus on the similarity relation between specific elements of the system, such as is emphasized in saying that Fortinbras is analogous to Laertes (as opposed to saying, oddly, that the “Hamlet-Fortinbras Complex” is analogous to the “Hamlet-Laertes Complex”). For this reason, I’d rather my account be able to accommodate either sort of emphasis, and to this end, it needs to be amended in the following way (the changes appear in boldface):

1. (a is analogous to b) *just in case*
2. (a and b are the same with respect to structure) *just in case*
3. **(either (there is some structure S, such that A and B both have S) or (there is some structure S, such that a and b both have S))** *just in case*
4. **(either (a belongs to a set A and b belongs to a set B such that A and B are denoted in common by some structure-predicate) or (a and b are denoted in common by some structure-predicate))** *just in case*
5. **(either (for some set of polyadic predicates denoting *n*-tuples of the elements of A, there is a set of polyadic predicates denoting *n*-tuples of elements of B, such that each element of A can be mapped 1-to-1 onto an element of B) or (for some set of polyadic predicates denoting *n*-tuples of the elements of a, there is a set of polyadic predicates denoting *n*-tuples of elements of b, such that each element of a can be mapped 1-to-1 onto an element of b) Or: a and b (or A and B) are structurally isomorphic.**

This change makes the definition indifferent as between analogies where analogues are parts of larger systems in virtue of which they *are* analogues and analogies where the analogues are the systems themselves. The Laertes/Fortinbras analogy is one of the first type. One of the second type is an analogy between *Hamlet* and *The Bad Sleep Well*.

II. Applying My Explication to Research on Analogy

In a survey of the progress that had been made in the twenty or so years since Connectionism began to contend with symbolic approaches to mental representation and structural accounts of analogy cognition were first proposed, two central figures in analogical modeling research, Hummel and Holyoak (1997), suggested fourteen criteria for evaluating models of analogy. Here is their table:

Empirical Phenomena for Evaluation of Models of Analogical Access and Mapping

I. Access and Its Relationship to Mapping

1. Semantic similarity has greater impact [on access] than in mapping
2. Isomorphism has less impact [on access] than in mapping
3. Close analog and schema easier to access than far analog
4. Access is competitive
5. Familiar analog accessed more readily

II. Analogical Mapping

6. Isomorphism
7. Semantic similarity
8. Pragmatic centrality
9. Multiple possible mappings for one analogy
10. Correct initial correspondence facilitates finding subsequent mappings
- *11. Difficulty finding mapping for "unnatural" analogy problems
- *12. Possible to map predicates with different numbers of arguments

III. Phylogenetic and Ontogenetic Change

13. Limited analogy ability exhibited by "language"-trained chimpanzees
14. Ability to process deeper and more complex analogies increases over childhood

* These criteria are considered plausible but lack direct empirical evidence.

Following my approach to similarity in the previous chapter, in the first section of the present chapter, I have provided a schematic account for an objective relation of analogy between two systems. In this sense, the account offered here is not meant to exhibit the empirical criteria identified by Hummel and Holyoak; the models they and other psychologists are talking about (even if such researchers are not always as clear as we might like on this matter) are psychological ones; they track psychological tendencies in employment, application, and articulation of analogies (or supposed analogies) not the difference between actual analogies and non-analogies. Nevertheless, especially when considering the list articulated by Hummel and Holyoak, the incorporation of an understanding of actual analogical

relations would prove most useful in these inquiries, because it could help clarify exactly what the models need to model. As it stands, the stated criteria reflect commitment to a number of errors regarding analogies, so a model that meets these criteria would, in this way, be of suspect value. Such a model would model general psychological tendencies regarding psychologists' ungrounded and flawed notion of analogy. I should think they would rather have a model modeling psychological tendencies regarding actual analogies. My explication can help them pose the questions in such a way as to make this latter *desideratum* attainable. I will now turn to discuss the stated "empirical phenomena" in turn, clarifying what they mean, pointing out ways in which they reflect erroneous beliefs about analogy, and finally suggesting possible alternative interpretations of the data they attempt to report that do not harbor the errors I identify.

As regards the search for psychological models of analogy, the field generally divides into two aspects: access, or the ability to call up a mental structure to function as a base domain to map onto a target domain; and mapping, the procedure that correlates elements of the target with elements of the base. In each of the next two sections I will discuss the "empirical phenomena" that fall under each of these headings.

A. Access and Its Relationship to Mapping

i. Semantic Similarity and Isomorphism

The first "empirical phenomenon" is that semantic similarity [between analogues] has greater impact [on access] than [it does] in mapping. What this

means is that semantic similarities between the target and the base analogue seem to play a greater role in a person accessing (from memory) a base structure when cued by a target analog than it does in mapping a base onto a target. That is, once a person has already got the two analogues to work with, they need not be semantically similar; but if a person only has one analogue and has to think of the other herself, then semantic similarity plays a greater role. This first criterion goes along with the second, which is really its counterpart: while semantic similarity is prominent in access, isomorphism is prominent in mapping. Semantically dissimilar isomorphs can be readily mapped, once they are available, but if only one analogue is cued, semantically similar analogues, which are non-isomorphic, are likely to be identified. Since these two criteria complement each other, I will discuss them together.

This pair of criteria, like so much else in the study of analogy, relies heavily on the notion of a *domain*. What is meant by the term 'semantically similar' is that the two analogues fall within the same domain. A paraphrase of the phenomenon in Hummel and Holyoak supports this conflation:

When analogs must be cued from long-term memory (rather than simply being stated as part of an explicit mapping problem), then cases from a domain similar to that of the cue are retrieved much more readily than cases from remote domains. For example, Keane (1986, Exp. 1) measured retrieval of a "convergence" analog to Duncker's (1945) radiation problem (for which the key solution is for a doctor to apply multiple low-intensity rays simultaneously to a stomach tumor from different directions). The source analog was studied 1-3 days prior to presentation of the target radiation problem. Keane found that 88 percent of participants retrieved a source analog from the same domain (a story about a surgeon treating a brain tumor), whereas only 12 percent retrieved a source from a remote domain (a story about a general capturing a fortress). (429)

While the notion of *semantic similarity* may be clarified by appeal to difference of domains, unless the latter is explained—which it is, nowhere in the literature on analogy, similarity, categorization, or learning—the former is not clarified. It might be countered that the use of ‘domain’ here is an everyday, commonsense one and that it is unfair to demand that it be explicated in such a context. Perhaps. But two considerations suggest otherwise. First, *domain* plays a theoretical role here, it is appealed to in a statement of empirical findings and thus is something that a theory (or model, in this case) is supposed to cover, explain, and reflect. Since this is the case, it seems difficult to defend the view that a “rough and ready” notion will do, given that, without some analysis, that notion might not track any salient empirical phenomena. Second, just taking the example at hand (Keane’s), and given that no analysis of *domains* has been provided, nothing seems to both put the two surgeon scenarios in the same domain and put the radiation and capturing general scenarios in different domains. Suppose the general flanked the fortress from all sides. This would put the general and the stomach tumor surgeon in the domain of “attackers from all sides”, a domain which excludes the brain tumor surgeon. The problem, thus, with this notion of *domains* (and its cohort *semantic similarity*) is that it gestures at nothing more than an ordinary notion of similarity, a thing which we have seen to be the cause of much trouble, at least when offered in the role of an explanation for something else, and especially as a heuristic device for finding out which sorts of things fit together and which don’t. What we see in this case, as well as many others that I have discussed up to this point is that appeals to similarity, or

here, sameness of a domain, don't function well without specification of some relative parameter.

A complementary argument holds against the second “empirical phenomenon”. Just as appeal to difference of domains does not separate the analogues that are easily accessed from those not so accessed (since, unless ‘same domain’ is given some substantive meaning other than just *similar* or “having to do with the same stuff”), the differences in mapping *vis-à-vis* access cannot be tracked along lines drawn by isomorphism, because all pairs are isomorphic. The idea intended, however, is something like the following: in cases where both analogues are known already, say, the two surgeons and the flanking general, the tendency is to compare the radiating surgeon with the general (instead of the “domain sharing” brain surgeon), because those two are isomorphic. The various rays of radiation coming from all sides map onto the various forces under the general's command, the tumor maps onto the fortress, the surgeon onto the general, and so forth. The problem with this, as I've suggested before, however, is that it doesn't distinguish the cases. The cases of the two surgeons are also isomorphic to each other, more directly, in fact: surgeon maps to surgeon, patient to patient, tumor to tumor, radiation to scalpel.

So it seems both of these first two “empirical phenomena” track distinctions without a difference. Yet, at the same time, they seem to gesture at something that is intuitively compelling. So, for all of my principled complaining and use of scare quotes, I by no means want to suggest that the research upon which these claims have been made is worthless or should be ignored. Rather, I want to find an

alternative way of articulating what is clearly aimed at in this research so that it need not be committed to the fake distinctions that it is relying on, as it stands.

These two findings seem to point to a tension between focus on more mundane sorts of similarities and deep structural ones. Under certain conditions, such as when an analogue case must be accessed from memory, the focus tends to fall on mundane similarities or *sameness of domain*; in conditions where the analogues are all given, and the structures are there on display, deeper structural similarities tend to be focused on. This being the case is not precluded by the fact that the alternatives not focused on in the first type of condition *also* share (many) domains with the cue analogue or the fact that the alternatives in the second type are *also* isomorphic to the target structure. However, in order to clearly articulate these data without recourse to an unqualified notion of *domain* or to wrong assumptions about isomorphism, some care must be taken. My accounts of similarity and isomorphism can help provide just the sort of care that is called for.

Let's look back at my first discussion above on the topic of domains and the cases of the general and the two surgeons. There I made the point that the radiation surgeon and the general share a domain, the domain of attackers from all sides. Now comparing that domain to the domain of surgical operation, it is obvious that the latter is generally more mundane than the former. 'Attackers from all sides' doesn't denominate a class of scenarios that is readily called upon in a very wide range of contexts, certainly not as readily as surgical operations is. This is a matter of degree, however; it is easy to imagine a person, say, a football coach, who thinks a lot about offensive strategy, who would readily identify the surgeon and the fellow general as

the counterpart to the radiation surgeon. One response would be to identify this as an aberrant case and the observed tendency to be subject to statistical error, another, for which the theory currently lacks the resources, would be to say that, for the imagined coach, the domain of “attackers from all sides” is a more salient one than the domain of surgical operations is. This alternative to the “it’s just statistical” approach also recommends itself, on the likely scenario where the observed regularities cease to hold when the subject pool draws from social groups other than those of the experimental designers and their undergraduate students.

In this way, making some modification to the objective notions of *domain* and *domain sharing*, a non-problematic restatement of the original “empirical phenomenon” is possible. Perhaps *domain sharing* does influence access in something like the way psychologists take themselves to have observed, but the domains to be shared must be relativized, to speaker or society or some other similar parameter or set of parameters. To recap, my counterargument has been based on the fact that any two analogues fall together within a single domain (as far as the notion of *domain* is understood through examples given by psychologists). Yet, this does not mean that everyone is aware of this fact or can make judgments based on it; and at any rate, even if some implicit knowledge of possibly obscure domains (like “attackers from all sides”) obtains for subjects unlikely to draw analogies based on them, there is surely a difference in psychological prominence between such domains and more obvious and mundane ones. Given these considerations, the “empirical phenomenon” can be recast as the following:

psychological prominence of the domains shared between analogues has greater impact on access than it does on mapping.

This alternative restatement clears away the theoretical confusions in the first “empirical phenomenon”. However, it generates an important side effect that may, from the standpoint of optimism for the progress of inquiry, appear rather deleterious. It introduces a new term ‘psychological prominence’. It necessitates studies of what makes a certain domain psychologically prominent. Fortunately, though, this is something that has already been underway, in the guise of studies into similarity judgments, although the two may feature different rubrics. The whole question of psychological prominence of various domains two items might commonly occupy is continuous with the study of similarity as I treated it in the previous chapter. Figuring out which respects of sameness are entrenched in a speaker or group’s judgments of similarity is not different from figuring out which *domains* are psychologically prominent for that speaker or group. If we know which respects of sameness tend to be employed in judgments of similarity, then we know which domains are likely to come to mind as ready candidates for a pair of analogues to occupy in common.

As a further move in clarifying my alternative, as well as to place it within the context I developed earlier in this section (the give-and-take between the first two stated “empirical phenomena”), we can see the difference aimed at by the first two phenomena as a difference between tendencies toward judgments of similarity and judgments of structural analogy under different conditions. While all pairs are both similar and structurally analogous to each other, it’s not true that there’s no

difference between similarity (sameness in a respect, or sharing of a *domain*) and structural or analogical similarity. Just as different *domains* or respects of sameness may draw natural focus for a given subject, given her background, experience and other qualities, general conditions may influence the type of relation on which she draws focus.

ii. Influence of Distance on Analogue and Schema

The third “empirical phenomenon” which Hummel and Holyoak identify as a necessary *explanandum* of an adequate model of psychological analogy is the notion that psychological access to a “remote” analogue is facilitated by certain learning conditions that encourage the induction of an abstract schema (1997: 429). What is meant by ‘remote’, here, is just that the analogues are not “semantically similar” or that they don’t share a domain. Since we have seen, in the previous section, that these distinctions are untenable, some modification of our description of the “phenomenon” must be made in order for it to count legitimately as evidence. However, the idea is that, while, under ordinary conditions, we may tend to focus on more superficial shared features of analogues, certain types of questioning and learning tend to help us focus on more structural and abstract features. So, going back to the general and the radiological surgeon again, instead of having to be given the general as an option in order to draw the structural analogy based on elements coming from all sides and converging on a target, we tend to be able to access these sorts of structural analogues on our own if we are first given sets of “semantically distant” analogues that share the type of structural similarity in question. So, I might

be able to come up with the general on my own as an analogue to the radiological surgeon if I've already been taught about the surgeon through comparison with other convergence scenarios.

Although there are problems with how this is formulated, because of the blind appeal to “distance” without a sound theoretical formulation of it, there is obviously something correct about the idea, so we should like to be able to recast it in a way that would allow for an unproblematic report; but, more than that, such a recasting is likely to clarify further research problems that as yet remain obscure because of the foundational problems created by the untenable appeal to “distance” that plays a role in the grounding of the theory.

The problem with *distance* is essentially the same as the problem with *semantic similarity* or *domain sharing*, which I discussed in the previous section. “Distant” analogues are supposed not to be semantically similar (or, alternatively, to share a domain). But, as we have seen from my earlier discussion, no pairs of analogues actually meet this criterion in a strict sense. However, analogues may share a domain (or fail to be “distant”) due to obscure and psychologically non-salient properties. If we import this notion into studies of learning that encourage induction of abstract structural relations (over surface features), then the data and their interpretation can be reformulated in the following way. Certain teaching and learning techniques induce a movement away from certain salient features (surface features) to other salient features (structural ones). In this way, the same sorts of studies can frame questions about the relation between teaching/learning techniques and sets of focused features for various subjects. This captures what is

intuitive about the original research without committing the projects to untenable theoretical commitments bound to an unsound notion of “distance”.

This modification, as well as the modifications suggested in section A.i., moreover, are not motivated merely by an overzealous philosophical pickiness. Rather, the rough and ready appeal in generalizations over evidence to concepts like “semantic similarity”, “domains”, and “distance” import the psychological prejudices of the researchers who formulate such generalizations into both research design and data interpretation. The fact that certain analogues, such as the radiological surgeon and the general with his surrounding forces, do not occur to researchers as similar (or, sharing a domain) does not entail that they actually are not similar, it only indicates that the respects of similarity in question (or the domains shared) are not psychologically salient to such researchers. By emphasizing these differences, based on the researchers’ own psychological tendencies, as theoretical differences (between *actual* distances, domains, or whatever), these researchers attribute their own psychological tendencies to others who may not share them. In this way, my overall critique so far of the “empirical phenomena” observed in relation to base-analog access suggests that some of the research questions have been posed and approached in something like a backward way. Instead of revealing differences in access to analogues based on distance, relative to teaching/learning techniques or various varieties of prior information, they seem rather to suggest that difference in teaching/learning techniques effect differences in which shared properties become actively focused on, perhaps relative to some prior set of psychologically salient properties for the subject or possibly groups to which she belongs.

iii.

The fourth “empirical phenomenon” observed in analogical access is competitiveness between possibly accessed analogues. This phenomenon broadly covers the fact that similarity (simple as well as structural) is not in itself sufficient for explaining cognition of analogy. This is not only consistent with the view I have presented in this chapter, it is *radically* implied by it. This is because, according to my view, the presence of similarity relations of various sorts (simple as well as structural) provides no distinguishing conditions, since such relations are present in every case. If such relations are always present, between any pair of items, it should not be any wonder that these in themselves cannot explain analogy cognition. So, in at least this much, the fourth phenomenon is one that needs no modification.

Though the fact that it has been included as a stated empirical phenomenon indicates the degree to which its radical overdetermination remains unappreciated by researchers. I take it to be a logical consequence of a sound theoretical understanding of analogy and, thus, a notion that becomes apparent prior to experimentation, not as a consequence of it.

iv.

The fifth “empirical phenomenon” is that familiar analogues are preferred (in access) to unfamiliar ones (even when the unfamiliar ones are “less similar” or “fit” less well than the familiar ones). According to Hummel and Holyoak (1997),

a particularly well-established example is the prevalent use of the person analog by children to make inferences about other animals and

plants (Inagaki & Hatano, 1987). It has also been shown that people understand new individuals by spontaneously relating them to significant others, such as a parent or close friend (Andersen, Glassman, Chen, & Cole, 1995).

These examples aren't at all surprising, and they seem quite intuitive. However, theoretically, the claim that they indicate or exemplify the use of familiar analogues over more similar though less familiar ones harbors problems that should by this point themselves be familiar. The notion of comparative similarity has not been provided with any sound theoretical clarification and seems merely to reflect the opinions of the researchers rather than a principled objective criterion.

However, the phenomenon (if we had a good description of it) exemplified by the examples comports well with the account I have developed in this chapter. My view has it that all things are similar (simply and structurally) in countless ways, but it, of course, does not attribute knowledge of these myriad ways to all (or even any) competent cognizers of analogy (not to mention of all competent researchers in cognitive psychology!). One characteristic of psychological familiarity is surely depth and/or breadth of understanding and awareness. Thus, it should not be at all surprising that familiar analogues are more prevalently accessed than unfamiliar ones, because familiar ones are the ones about which more is very probably understood. If I don't know that two items share a given property (or structure), I don't have much of a chance of being able to correlate those items based on that property. And If I don't know much about the items in question, then I don't have much of a chance of knowing which properties they have or which structures they instantiate.

Furthermore, this sort of statement, as a confirmed “empirical phenomenon”, brooks any principled response from empirical psychology that, since that discipline’s concern is psychological analogy (or, similarity, as in the previous chapter), my objective schema is irrelevant to its research goals. That is, examples of this sort show that, even though the research aims to understand how people process analogies psychologically (and not to develop an objective theory of analogy, in fact), the beliefs and intuitions about actual analogy held by the researchers influence their experimental design and interpretation in substantive ways. This is because, without an appeal from the research designers to the fact that some analogues may “fit better” with, say, a palm tree, than a person does, no sense can be made out of the notion that familiarity must be added to fit to described conditions of analogue access. From the point of view of the children described in the excerpt from Hummel and Holyoak, a person may well be the “best fit” *available* for the palm tree. Thus, in order to describe the data in the way that they do, they must rely on their own judgments about what pairs are good examples of analogues, or which are “good fits”. However, as I’ve argued throughout this chapter, these judgments are not well founded. My account is offered as a means to correct this situation; where judgments of analogical relations are needed in formulating research questions, designing experiments and interpreting the collected data, my account can be relied upon as an objective framework in the place of the unsupported (and often erroneous) judgments of researchers in the field.

This concludes my discussion of the “empirical phenomena” that researchers have observed in relation to psychological access to base analogues for target

analogues cued in various ways. I will now turn to discuss the phenomena relating directly to the process of mapping from a base analogue onto a target analogue.

B. Mapping

i.

One clear criterion that has been observed to influence mapping base analogues onto targets in empirical studies of analogy cognition is isomorphism between the analogues. Since all pairs are isomorphic, this, on its own, hardly narrows the field or bears much explanatory fruit. However, many studies go beyond the mere recognition that isomorphic relations exist to further investigate which such relations are used in actual cases of mapping and (thus, I would argue, which structures instantiated by the relative analogues are more and less psychologically salient).

While I think that any satisfactory account of analogy (in either psychology or philosophy) will incorporate appeal to isomorphism and that psychologists are right to maintain that their models must have a place for isomorphism, some of the consequences that have been “observed” are problematical. For example, Gentner and Toupin (1986) show data purported to confirm that isomorphism can underwrite mappings even when they conflict with object similarity. Such a statement involves two notable problems that I’ve raised a number of times already. There is no free-floating object *dissimilarity*, so there aren’t genuine conflicts of object similarity. Again, what would be needed here is some discussion of ranges of

psychological salience, comparing single one-place predicates (representing “object similarity”) with sets of relations (representing isomorphic structures).

But suppose that, instead of talking about similarity and isomorphism in nonspecific ways, we knew something like the respects in question. Even with parameters of that sort set and specified, a second problem arises. The second problem is the fact that the difference between the sharing of a structure and being *object similar* (or, what has been variously called ‘simple’ or ‘semantic’ similarity) is syntactic, not semantic. The same facts about an object or pair of objects may be schematized using semantically dense one-place predicates, such as ‘___ loves Romeo’, or they may be schematized in a way that shows relational structure (using a more articulate predicate, such as ‘___ loves ...’). This creates problems for the sort of data interpretation, such as Gentner and Toupin’s (1986), that suggests isomorphism, under certain conditions, weighs against object similarity, because it illustrates that whether a given pair of objects (or systems) is isomorphic or object similar *even in a given respect* isn’t something that has a clear, non-relative answer. Of course, given, say, the love affair between Juliet and Romeo and the familial relation between Hamlet and Claudius, we are likely to understand each as a relation between elements. In the first case, ‘___ loves ...’ with either ⟨romeo, juliet⟩, ⟨juliet, romeo⟩ or both in the extension. In the second case, ‘___ is uncle of ...’, is ‘___ stepfather of ...’, ‘___ is nephew of ...’, and so forth, with the appropriate extensions. Regimenting the structure of each situation in some of the ways suggested leads to obvious isomorphic mappings; Juliet onto Hamlet, Romeo onto Claudius, ‘loves’ onto ‘nephew of’, for example. But, since the formal characteristics of these relations are

underdetermined by the relative situations (the loving and the nephewhood) only after the regimentation has taken place, it cannot be said that these situations are *fundamentally* relational. Rather, they are relational relative to some procedure for analyzing them into significant parts. According to the procedure I've sketched, Romeo, Juliet, Hamlet and Claudius all turn out to be significant parts, and thus in a regimented statement of the situational facts, these parts stand in certain relations to one another. I could have, instead, taken only, say, Juliet and Hamlet to be significant parts (in terms of the situations' ontology) and thus described them using one-place predicates such as '___ loves Romeo' and '___ is Claudius' nephew'. It might be objected, though, that this route precludes what the relational route doesn't; that is, it undermines the basis for claiming similarity of any sort. When we analyzed the situations into a relational structure, it gave us the possibility of mapping the elements onto each other in light of the relations that held between them. Here, there are not such relations, just one-place predicates, and, at least in the account of similarity I've given (and the various approaches taken in cognitive psychology), two things must at least share a property (or be denoted in common by a predicate) in order to count as similar.

But this sort of reply, again, reflects the confusion of taking certain properties of the situation as more fundamentally *given* than others. It may be true that Juliet isn't like Hamlet with regard to whom each loves or to whom each are related. But they are alike in other ways, even in given stated respects. Juliet and Hamlet are alike in loving someone whom it is socially unacceptable to love, and they are alike in being related to someone who willfully harms someone they love. Since this is the

case, it cannot be enough to compare cases as between isomorphism and “object similarity”, at least if the decision about which type of similarity obtains is left up to the unreflective judgments of the experimental designers. This does not mean that it would be impossible to tell, under certain lines of questioning which type of similarity (isomorphic or object) is *in fact* relied on in specific cases of subjective judgment; rather, it suggests that, starting with the experimenters’ unreflective beliefs about these matters leaves important distinctions about isomorphism and object similarity beneath the radar of the experimental data. The fact that a researcher regards a certain object pair as isomorphic does not entail that subjects who judge that pair as similar do so because of the isomorphism (and in spite of object similarity), since, regardless of the perceptions of the researchers, the pair of items in question (whatever they may be) *are* object similar. In order to show that it is the isomorphism that the researchers have in mind and not the (sometimes weird) object similarities that underwrite subjective judgments of similarity, further questions must be asked. And this is, again, not a critique that should be taken as a mandate to throw out great and extensive research efforts already begun and still underway; it is offered as a theoretical development that should highlight ways in which that research could focus on distinctions it already employs but to which its outcomes are often indifferent.

Just as, in the beginning of this section (B.i.), I agreed with Hummel and Holyoak (as representatives of the mainstream of analogy research) that isomorphism is a relevant phenomenon for analogical mapping. The second phenomenon they identify, semantic similarity, is also in some sense relevant. But,

as I have argued in this and previous sections, without some relativization, appeals to semantic similarity do little more than show where the pretheoretic judgments of the researchers show up unreflectively in their interpretations of their own data. Thus, the body of what I've said here about isomorphism applies as well to semantic similarity, so I will move on to discuss other "empirical phenomena" without further discussion of semantic similarity.

ii.

The remaining observed phenomena all involve ways in which analogue pairs underdetermine mapping relations. And, in a way, this indicates internal inconsistencies in the theoretical structure of analogy research in cognitive psychology as well as suggests that, at least to some degree, the main thrust of my position is nascent in that field, even if it is not incorporated explicitly or consistently. Thus, it is well understood that neither isomorphism nor "semantic similarity" nor the two together can fully explain analogy cognition, because these properties alone (and the conjunction of them) describe many, many more analogies than can ever be understood or intended in a specific case. Or, more emphatically, it provides no information to say that isomorphism, semantic similarity, or both are necessary conditions for analogy cognition, because these conditions never fail to obtain; and it is wrong (as psychologists realize) to say that either property (or their conjunction) is sufficient, because not all analogies are recognized in any particular case, even though these properties always obtain. For any given pair of analogues, even for someone who understands them *as analogues* and can describe complex

mappings of elements from one analogue onto elements of the other so that all sorts of interesting structural features are preserved, there will still be countless other mappings, both bolstered by and at variance with certain obvious semantic similarities. Thus, some studies focus on features other than isomorphism and semantic similarity to explain analogy cognition; these efforts represent the remaining “empirical phenomena” I will discuss: multiple possible mappings for one analogy, and correct initial correspondence facilitates finding subsequent mappings.

Hummel and Holyoak report one empirical finding in studies of analogy cognition to be that there are multiple possible mappings for one analogy. While this is true, it should not be taken as following from empirical evidence in an ordinary sense of empirical confirmation, rather it follows from a correct theory of analogy, generally. In this sense, the stated “empirical phenomenon” could be refashioned to state that it has been observed that subjects sometimes recognize that there is more than one correct mapping, or that various subjects sometimes map the same analogies differently.

While the stated phenomenon is veridical, the interpretation of that phenomenon in the literature belies misunderstanding of the phenomenon. For example, Hummel and Holyoak liken the phenomenon to a Necker cube, saying “people typically arrive at one interpretation of an ambiguous analogy (although they may be able to shift from one interpretation to another)” (1997: 430). The comparison to the Necker cube and the designation of some analogies as “ambiguous” suggests that, for the examples they have in mind, there is a certain limited number of ultimately *correct* ways in which to interpret the analogy in

question. There may be some sense to be given to correctness of mapping once it is made clear what problems are to be solved or inferences to be made (or even just which structures are to be compared), but on similarity alone (either structural, “semantic”, or both) no sense has been (nor, I have argued, could be) given to the notion of a correct mapping beyond the universally met constraint of isomorphism. So, instead of using the notion that there are multiple possible mappings in the sense in which it is true, as a foundational assumption that influences the posing of research questions and formulation of methodologies, it seems as though this fact indicates a detrimental lack of awareness as to the full extent of the marked overabundance of mapping possibilities.

The notion of correctness in the drawing of analogies also shows up in the next “empirical phenomenon”: correct initial correspondence facilitates finding subsequent mappings. This phenomenon is based on the work of Holyoak and Thagard’s (1989) and, especially, Keane (1995). In these examples, however, the notion of correctness isn’t so much the idea of drawing analogies correctly in general; rather, it is correctly mapping structurally described sets of facts onto other sets. For example, one of the mapping problems Keane uses is found in the following table (1995):

Table 1: Examples of mapping problems used in Experiment 1

<i>Singletons-Aligned</i>		<i>Singletons-Crossed</i>	
A	B	A	B
Joe motivates Steven.*	Lisa hugs Jenny.*	Mark is beside Ronan.	Lisa hugs Jenny.*
Mark is beside Ronan.	Laura employs Ruth.	Mark motivates Ronan.	Laura employs Ruth.
Mark motivates Ronan.	Laura hugs Ruth.	Conor is beside Paul.	Laura hugs Ruth.
Conor is beside Paul.	Mary sees Ali.	Conor fears Paul.	Mary sees Ali.
Conor fears Paul.	Mary employs Ali.	Joe motivates Steven.*	Mary employs Ali.

* indicates the singleton

This table was used in studies to show that the order of presentation influences the incidence of error in mapping problems. If the singletons (the only arguments to occur only once, Steven and Jenny) are presented aligned, mapping is completed correctly more often than when they are presented nonaligned. This example indicates an important distinction that has not yet become explicit in my discussion; that is the distinction between analogies between things (or sets of them) and between descriptions of them. Much of my main line of discussion has been based on the overabundance of analogies between things; that overabundance is due to the fact that the world will always provide facts to ground mappings from a base to a target. However, this may not be the case once the analogues get “summed up” in a neat description. This can be seen in the table above; one way of excising lots of live alternative options for analogical mapping is to describe the base and the domain in a way the precludes such alternatives. While there may be all sorts of isomorphic mappings from Steve, Ronan and Paul onto Jenny, Ruth, and Ali (whoever they are), if all we know about these people is stated in the columns above, then there seems to be only one correct (isomorphic) mapping of one onto the other.

However, I would say that even in such an example, where our knowledge of

the situation is so much more impoverished than it ever would be in a case where direct observations are made, alternative isomorphic (and therefore, by this standard, *correct*) mappings can be identified. For example, to the original list of descriptions,

A	B
Joe motivates Steven.*	Lisa hugs Jenny.*
Mark is beside Ronan.	Laura employs Ruth.
Mark motivates Ronan.	Laura hugs Ruth.
Conor is beside Paul.	Mary sees Ali.
Conor fears Paul.	Mary employs Ali.

we can make many easily justifiable inferences, so that the determined state of affairs as described admits of multiple correct mappings. For ‘motivates’, ‘fears’, ‘hugs’, ‘employs’, and ‘sees’, we can add the corresponding passive restatements of all sentences with those terms in them. So, in addition to ‘Joe motivates Steven’, *A* would also have ‘Steven is motivated by Joe’, and so forth. Adding only some of these restatements undermines the single “correct” isomorphic mapping. However, it might be objected on some spurious epistemic grounds that all restatements must be added: something like adding only some presents an incomplete picture of the situation. Suppose this were granted; new problems then arise; ‘is beside’ is a copular construction and doesn’t admit of active and passive voice. So, while ‘is beside’ in the original formulation maps onto ‘employs’; there is no passive for ‘is beside’ to correspond to ‘is employed by’. Adding all of the passives itself undermines the single “correct” mapping. Further alternatives are introduced if we recognize that ‘is beside’ is symmetric, and we can thus add ‘Ronan is beside Mark’ and ‘Conor is beside Paul’. This changes considerably the range of possible mappings. And excluding these as legitimate options entails rejecting the idea that if

I know Mark is beside Ronan I also know that Ronan is beside Mark. And these additions are just the most obvious and immediate ones focusing specifically on the relations in the lists. We could also help ourselves to all sorts of other facts about Joe, Lisa and company, such as about their probable genders, about where their names appear on lists used in psychology experiments, and so on.

To summarize this last argument, then, one objection that might be brought against my whole line of argument is that my critique of psychological methodology and data interpretation glosses a distinction between analogical relationships between things (or sets of them), which *do* admit of radically different and many interpretations underwritten by isomorphic mappings, and analogical relationships between *descriptions* of things, which do not admit of variant mappings. I have shown, using the example from Keane, that even when it looks like the information about the objects in question is highly opaque, it should always be possible to extrapolate further information so as to admit of various mappings. I should think that the only time a little ingenuity and creative attention wouldn't reveal alternatives of this sort is in cases where structures are described formally (using, schematic predicate letters and a set universe of objects, say) so that the semantic content is removed and contact with worldly facts is made impossible.

C. Summary to Section II.

In this part of the chapter, I have discussed a series of interpretations of subjective research data that have been offered as “empirical phenomena” which any adequate psychological model of analogy will explain or predict. I have

subjected each of these interpretations to a critique, under the auspices of my own explication of analogy. To this end, I have attempted to replace intuitive and, in some previously unspecified sense, correct findings with reformulations that not only avoid the errors I identify but suggest further focal points for refined research goals. At this point, I will now discuss some features of the hierarchical model characteristic of most accounts of structure in analogical research.

III. Relations between Relations; an Attempt at Hierarchy within a Structure.

One of the problems still outstanding is that of separating useful or “good” analogies from other analogies. This problem can be seen, variously, as one for the psychology of analogies, the problem of differentiating the analogies or sorts of analogies we *tend* to focus on from those we tend not to, as well as one for the logic of analogies, the problem of differentiating the types of analogies we *ought* to focus on from those we would do well to ignore in making comparative arguments.

Insofar as we are good reasoners in our use of analogies, these will coincide, insofar as we aren’t, they won’t. In general, however, very little headway has been made in understanding the logic of analogies. My own view of the matter is that analogies are not usefully logical, in the sense of providing independent, truth-preserving support for outside claims. They might provide inductive support in certain instances, but only of a kind distinct from ordinary inductive arguments and only as ancillaries to such arguments. In this sense, analogy, within heuristic epistemology, functions most effectively as a tool for creative thought and hypothesis formulation rather than a particularized variety of inductive confirmation (as it has been treated by

some philosophers working on confirmation theory). These issues will be addressed in my next and final chapter.

A. Characterizing Higher-Order Relations.

In this section, I will focus on the psychological notion of a good analogy. Specifically, most psychological models differentiate the good analogies by identifying not just a shared structure between base and target systems but also according to how “deep” these shared structures are. The image is of a hierarchical ordering of nodes. So that “good” analogies are those where not only a relational structure between elements is shared but a relational structure between relations themselves is shared as well. This idea goes as far back as Gentner and Gentner’s (1983) foundational paper that has initiated a whole generation of research on the psychology of analogy that focuses on shared structures rather than shared properties. In that paper, Gentner and Gentner identify “systematicity” as a quality of structures (in addition to relational mapping) that makes them good candidates for use in analogical reasoning; by this they mean “predicates are more likely to be imported [for use in analogical reasoning] if they belong to a system of coherent, mutually constraining relationships, the others of which map onto the target. These interconnections among predicates are explicitly structurally represented by higher-order relations between those predicates”(1983: 104). If what has been one of my persistent focuses throughout this and the prior two chapters has a glimmer of recognition anywhere in the psychological literature, it is here. That is, structure-mapping can’t alone ground an account of analogy because its possibility is

ubiquitous. Thus, some other constraint must be added. This notion isn't fully recognized in the addition of "systematicity" as a second criterion, rather what is recognized is that, in certain good cases of analogy, there are relations in the base system, such as, say, the temperature of the sun compared to that of the planets in the solar system/atom analogy, that don't have obvious correlates in the target system. However, the criterion of "systematicity" hasn't been given a sufficient characterization; specifically, the notion of relations forming an interconnected network due to the hierarchical ordering effected by higher-order relations does at least one of two bad things: it either commits a use/mention error or harbors confusions about what relations are. In the remainder of this section, I will explain this dilemma and suggest a way in which the notion of systematicity aimed at could be given sense.

The idea that systematicity (and its attendant higher-order relations) attempts to capture is that there are some relations between elements in one system that are just *random*, and the fact that there isn't a good correspondent relation in another system doesn't undermine a possible analogy in the sense underwritten by a shared structure; it just means that that random (non-systematic) relation isn't a part of the structure to be mapped. Recalling again the solar system/atom analogy, a couple of relations that make that analogy compelling are that, in both systems (solar system and atom), the less massy objects (planets, electrons) revolve around the more massy ones (sun, nucleus); and the significance of this correlation isn't undone by the fact that lots of relations between the sun and planets don't hold between the nucleus and the electrons and *vice versa*. The relations that *do* hold are thought to

maintain significance in spite of numerous *disanalogies*, because they are systematic; or – what this systematicity is taken to be –there are higher-order relations between these relations. The following diagram exemplifies the hierarchical structure this model aims to incorporate:

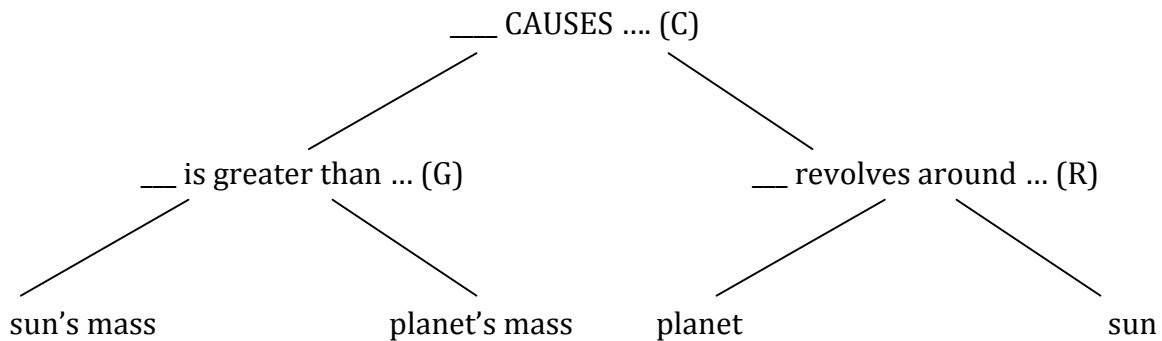


Figure 1

The analogous model for the atom can be obtained by replacing instances of ‘sun’ with ‘nucleus’ and instances of ‘planet’ with ‘electron’. Under the proposed analysis, the relations ‘G’ and ‘R’ are taken to be systematic, by virtue of the fact that there is a higher-order relation, ‘C’, that relates them. The fact that the sun’s mass is greater than the planet’s (or that the nucleus’ is greater than the electron’s) isn’t just a random fact *vis-à-vis* the revolution of the one around the other; These sets of facts are related systematically, and the mark that identifies such systematicity is taken to be the higher-order relation ‘C’ that relates the two lower-order relations, ‘G’ and ‘R’.

That all sounds attractive. However, it’s also rather a bit hasty, and it comes apart under a very little scrutiny. The one culprit here is the equivocal use of the term ‘relation’ in the characterization of higher-order relations as “relations between relations”. As it stands, and since it hasn’t been outfitted with a technical

specification, 'relation' is ambiguous (at least) as between a function that maps pluralities of objects onto truth-values and the fact that, say, a certain plurality *in fact* maps onto true. Or, to use an example, 'relation' is ambiguous between '___ is greater ...', which is a relation, and the fact that the mass of the sun is greater than the mass of a planet, which reports a relation.

This ambiguity might simply be a bit of unimportant carelessness that needs to be cleared away, if settling the sense of 'relation' in either direction would yield a coherent account. The problem, however, is that characterization of a higher-order relation as a relation between relations actually *relies upon* this ambiguity. In this characterization the 'relation' in 'higher-order relations' refers to the mapping function, in this example, '___ causes', and the 'relations' in 'relations between relations' refer to the facts that certain functions map to true, not to the functions themselves. 'Relation' must be used equivocally, or the description of systematicity doesn't make sense. This can be shown easily by the use of two instantiations of a non-equivocal reading of the account. Taking 'relation' non-equivocally in the functional sense, a saturated higher-order relation would look like the following:

a) '___ is greater than' causes '___ revolves around'

If a) makes sense at all, it has little chance of being true. It does indicate, at least, '___ causes ...' as a "relation between relations", in a consistent sense, however. In order to get clearer on the problem, to a), we might compare, say, '___ is to the left of on page 171', as a relation between relations in the non-equivocal, functional sense of 'relation'. Saturated similarly, we get the following:

b) '___ is greater than' is to the left of '___ revolves around' on page 171.

At first blush, it looks like there is a fair possibility that b), unlike a), is actually true. Furthermore, b) seems to represent a fair example of a relation holding (or possibly failing to hold) between a pair of relations, in the functional sense of ‘relation’. However, if we take seriously the notion of a relation as a *function*, a mapping between object sets and truth-values, it becomes clear that b) is not true either. The two functions *denoted* by ‘___ is greater than ...’ and ‘___ revolves around ...’, which are themselves abstract objects, don’t appear anywhere on page 171; rather, inscriptions denoting them appear on page 171. So, while we can make sense of b) as indicating a “relation between relations”, it still doesn’t provide an example of such a relation actually holding. For that, c) provides an example:

- c) An inscription denoting the relation ‘___ is greater than ...’ is to the left of an inscription denoting ‘___ revolves around ...’ on page 171.

This unwieldy relation exemplifies one that not only holds between relations, but holds *only* between relations. And that is, I should think, what would be wanted out of a characterization of higher-order relations as those holding between relations. However, once we actually get a look at an example that fits the proposed description (a “relation between relations”), we see that it has no chance at all of doing the work intended. This kind of “relation between relations” isn’t at all likely to differentiate the empirically interesting structures from the random ones: there seems to be nothing “systematic” about it, in the intended sense. The “higher-order” relation in c), for example, is just about as random a relation as one could hope to find. And if that blanket assertion weren’t enough to close the case, it will do to point out that a similar relation could be easily cooked up that would hold between say, ‘___ is greater than ...’ and ‘___ is hotter than ...’, the latter of which, as noted above,

was initially described by Gentner and Gentner (1983) as non-systematic because of the lack of higher-order relations relating it to other relations holding between the elements of the system.

Furthermore, if we attempt to read the characterization of higher-order relations as relations between relations non-equivocally in the other direction, that is, not in the functional sense of ‘relation’, but in the sense of veridically saturated functions, we make even less headway. ‘Relation’ in ‘higher-order relation’ *must* mean ‘relation’ in the functional sense. To say that something is a “relation between relations” is just to say that it is a function taking relations as arguments.

What is clearly wanted out of the “relations between relations” is exemplified by d):

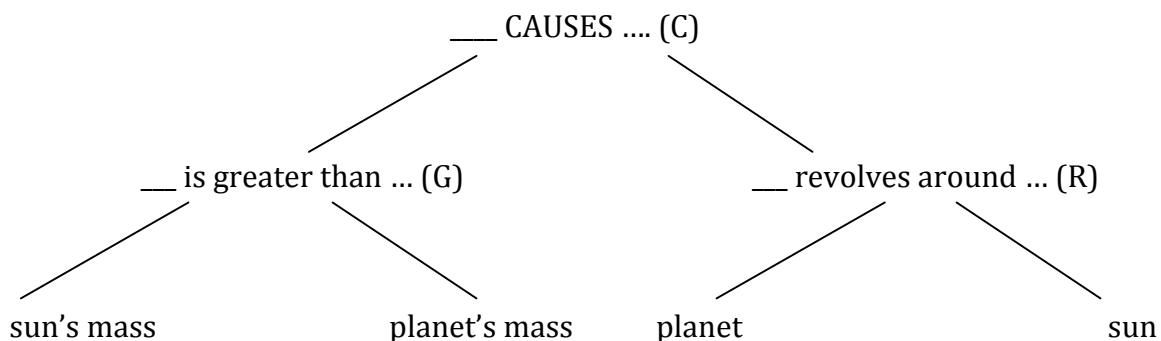
- d) (the sun’s mass is greater than the planet’s mass) causes (the planet revolves around the sun)

I have argued in the foregoing that this does not represent a “relation between relations” and, furthermore, that genuine relations between relations are weird and unable to ground systematicity in the way aimed at.

Other authors cash out the higher/lower-order difference in terms of propositions as arguments rather than in terms of relations as arguments. This seems to get closer to describing the example of d), but harbors the same problem in a new form. If we say with Wilson *et al.* (2001), for example, that higher-order propositions take chunked propositions (instead of objects) as arguments, then we’re stuck again not really describing what is wanted. Taking d) seriously, as a (higher-order?) proposition, weird things appear. It’s not, after all, the proposition ‘the sun’s mass is greater than the planet’s mass’ that causes *the proposition* ‘the

planet revolves around the sun', which is what would have to be the case for '___ causes ...' to take propositions as arguments (and map to true under this interpretation). In this instance, it's not the propositions that are causally related, but rather something like "the facts" they report. The fact *that* the sun's mass is greater than the planet's *causes* the fact that the planet revolves around the sun. At this point the shallows of metaphysics have already become too deep for me, and I should say that I'm not ontologically committed to either facts or to propositions²⁸. I've used that kind of terminology only to show that if one accepts propositions (as anyone who's theory draws on a difference between higher- and lower-order propositions does), then identifying the higher-order propositions as those that take propositions as arguments fails to meet the intended goal. What is meant is not that the propositions are the arguments but rather that what such propositions report, or name, are; some people have called these referents 'facts'.

At this point, it will do to take stock a bit before I move on to criticize this way of specifying what makes a certain structure a good one for analogical transfer. We can do this by looking back once more to the hierarchically represented structure I displayed in figure 1 above. Here it is again:



²⁸ At least insofar as propositions are supposed to be something distinct from (but somehow related to) sentences.

Under the modified analysis I've offered, 'C' can be described as a higher-order relation, because its arguments aren't ordinary objects but are whatever propositions name. Each node of the diagram contains an inscription, and correctly interpreting the diagram involves understanding each inscription as being *used* instead of mentioned. The problem with the characterization of the higher-order relation as taking propositions as arguments was the problem of reading the middle nodes (the first order relations) as being used with respect to the lower nodes and mentioned with respect to the higher ones. This problem is, therefore, analogous to the problems I discussed with characterizing higher-order relations as relations between relations.

B. Critique of Hierarchical Ordering as a Distinguishing Criterion for Good Structures

The problem with identifying hierarchical ordering as a mark that distinguishes the types of structures that tend to be focused on successfully in analogical reasoning problems is that this mark fails to draw any distinction that cuts along lines of pretheoretical judgment about systematicity. For any set of objects, there will be ways of specifying the instantiated structure so that it has this sort of hierarchical ordering.

Looking back to Gentner and Gentner's example, cited above, a relational predicate taken *not* to be part of the relevant structure for analogical transfer from the solar system to the atom is '___ is hotter than ...'. However, as I suggested above,

it is entirely possible to cook up a relation that holds between the fact that the sun is hotter than a planet and the fact that, say, the sun has more mass than a planet. Thus, according to the criterion of hierarchical ordering set forth, ‘___ is hotter than ...’, taken to be a prime example of a bad candidate for structural transfer, would be included in that structure to be transferred. At this point, the defender of the criterion might object that although it may be possible to cook up a weird relation to hold between the stated facts about the solar system, that would only fill in half the picture, because the structure instantiated by the atom would have to have a relation corresponding to ‘___ is hotter than ...’ *and* one corresponding to the “cooked up” higher-order relation. But a little ingenuity would easily satisfy this constraint. In 1897, J.J. Thompson discovered the existence of electrons; until that time, atoms were thought to be indivisible. But Thompson thought atoms were composed of evenly distributed electrons; it wasn’t until 1909 that Rutherford first suggested that most of an atom’s mass is concentrated in the nucleus. Thus, just as the sun is hotter than a planet, the nucleus was discovered later than the electron. And that fact that the nucleus was discovered later than the electron can be related (higher-order) to the fact that the nucleus is massier than the electron. Thus, we can generate a pair of analogous, hierarchical structures that are non-systematic:

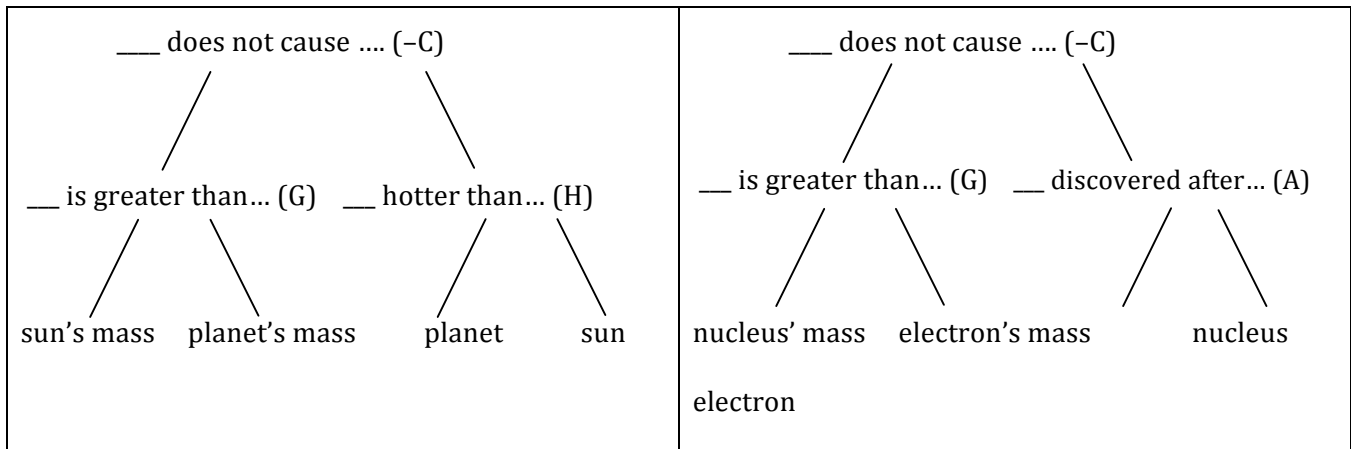


Figure 2

In either of the two stated senses, and applying the interpretation I have suggested, the only one that makes sense of “higher-order relations” as characterized, this example demonstrates that the “deepness” achieved by a hierarchical structure, typified by higher-order relations, does not capture what was wanted out of systematicity. This example meets the stated criteria, and, thus, according to those criteria would have to be deemed systematic, although it also incorporates the relation ‘___ is hotter than ...’, which, by Gentner and Gentner’s own lights, exemplifies the sort of relation that would rightly be excluded from a systematic structure, and, thus, would be deemed inappropriate for analogical transfer.

C. How to Proceed

If the critique of systematicity I have offered is valid, it altogether undermines the foundation of much research into the cognition of analogy. What was wanted was for systematic structures to be the type of structures that are

“good” for analogical transfer, but it turns out that structures clearly not good (and admitted to be) also fit the stated criterion of systematicity. I think this is, at bottom, a problem of epistemology; the attempt offers a formal constraint as a mark of an empirical difference. Since systematicity is understood or taken to be a condition that can be characterized by formal properties of a structure – that the structure is hierarchical, and this hierarchy obtains as a consequence of certain types of interrelations in the structure – there will always be facts enough to satisfy that formal condition. Again, just as in the many other cases I’ve discussed before, much research design and interpretation seems to hinge on the fact that certain truths that don’t occur to researchers, such as the fact that the relative masses of the sun and a planet *do* relate to their relative temperatures, in certain *odd* ways. From this lack of recognition, it is inferred that such truths don’t obtain, and the conclusion of that inference becomes incorporated, problematically, into the basis of the model. Looked at in this way, the criterion of systematicity offered to differentiate structures devolves from an objective distinction tracking the world and psychological tendencies generally to one that merely tracks the researchers’ pretheoretical judgments about what counts as systematic and what doesn’t.

The rejection of systematicity as characterized suggests some possible avenues of recourse. There *does* seem to be something right about identifying the structure (and its correlate for the atom) exemplified in Figure 1 as systematic, in contrast to the structures exemplified by Figure 2. The first set seems to have something natural and “deep” in common and the second set seems to be just randomly correlated. What remains missing is a way of characterizing this

difference. I suggest comparing the two pairs in terms of what *roles* they might play, relatively, in reasoning, learning, creativity, and cognition. The structures exemplified in Figure 1 (and its partner for the atom) could reasonably be used in teaching atomic theory and even in the development of atomic theory in the first place. These structures are underwritten by more general empirical commitments, generalized correlations between empirical phenomena; the structures in Figure 2, however, have no underlying empirical basis according to which they either cohere individually or relate to each other. Using such examples in a teaching/learning context, or a context of creative theoretical thought is likely to end in confusion, at best, and possibly lead to serious error. These structures don't "fit well" into an overall picture of empirical reality as we commonly understand it.

A rejoinder to this suggestion might be that it relies on vague and often transitory notions, such as "fit" and "an overall picture of empirical reality". According to the case I have developed against the psychologist's notion of systematicity, this is as it should be, however. The problem with that notion was that it mistook an empirical distinction for a formal one. Since analogical reasoning is a type of inductive reasoning, there should be no formal distinction between those analogies that count as worthwhile and those that count as worthless – or between systematic and non-systematic varieties – this distinction should be something that develops and changes as we, either globally or individually, learn more and more about the world and the way that it works. In this sense, the psychologist's problem of distinguishing systematic analogies from non-systematic ones is difficult to

distinguish from the philosopher's of evaluating the use of analogy in logic and argumentation.

IV. Conclusion

In this Chapter, I have criticized the notion of analogy and used my own, previously articulated account of similarity to develop a schematic device for explicating analogies in particular cases. Using this schematic, I have criticized and commented upon the main line of empirical inquiry regarding analogy cognition in cognitive psychology. The main thrust of this critique has been to show that it is often the case that psychological research draws distinctions without a difference, contrasting simple similarity to analogical similarity, for example, when both relations obtain in the particular cases. Where the observations and research questions have seemed, nevertheless, to gesture at truth and suggest a fruitful path of inquiry, I have suggested ways in which, under my account of analogy, the questions posed and interpretations offered might be modified so as to avoid further error and confusion. A fundamental culprit identified in all of these arguments is the inclusion of the pretheoretical judgments of the researchers in the formulation of research questions and data interpretation, a problem particularly worrisome when these judgments are subject to clear counterexamples, which I have offered. In the final section of the chapter, I have discussed one of the basic distinctions drawn in virtually all analogical research, that between systematic and non-systematic structures. I have shown first that, as articulated, this distinction lacks sense, and, second, if we restate it to preserve the original insight in a coherent

way, it fails to cut along the lines suggested. That is, those structures meeting the stated criterion of systematicity need not thereby be meaningfully systematic. I have concluded with some tentative remarks about what is wanted out of systematicity, which relate directly to the topic of the next and final chapter.

CHAPTER 5
ANALOGIES IN PHILOSOPHY AND UNCHARTED SPACES:
RESPECTIVISM AND STRUCTURE-PREDICATES

I. Getting Out of a Head Detail

From early on in Chapter 1 and variously in Chapter 2, I have maintained that one reason for recommending explication as a philosophical methodology or an activity for philosophers to pursue is that explications themselves, while they may end up departing from the ordinary concepts from which they arise and may end up dissolving rather than resolving the philosophical puzzles which motivate them, can serve to create new tools for investigation and analysis that were unavailable without the explications. My notions of sameness in a respect and of structure-predicates offer promise of this sort. In order to fill out these possibilities, I will review my explications of similarity and analogy, reviewing the idea of sameness in a respect and pointing out how the notion of a structure-predicate comes out of the unified accounts. Before discussing how these concepts hold such promise beyond their applicability to method in cognitive psychology, however, I want to discuss an example of this sort of promise paradigmatically actualized. In Chapter 2, I mention the operators and quantifiers of modern formal logic as examples of explicated concepts from ordinary language²⁹. In the context of this discussion, I indicated how these explicated terms depart from ordinary usage and thereby avoid some of the vagueness and imprecision of that usage (and, thus, dissolve some of the puzzles consequent of these qualities). More is gained from these departures, however, than

²⁹ P. 65

clarification and dissolution of elementary puzzles.

A famous example of the advance in knowledge made possible by the explications of modern symbolic logic is the dissolution of the various conundrums surrounding the argument of the horse's head: Horses are animals, therefore every horse's head is an animal's head (DeMorgan 1874: 114).³⁰ Various philosophers and logicians from Aristotle until the mid-19th century, especially Aristotle's Scholastic acolytes, have engaged in all manner of contortionist reasoning to explain how an argument like this fails to be valid or how it relies on extra-logical content³¹, since it cannot be formulated into the Aristotelian syllogistic formulae for logical validity. The formulation, after Frege, of logical operators as well as polyadic predicates whose variables could be bound by (multiple) quantifiers made it possible for the first time to prove the conclusion of the horse's head argument from the single premise using only universal and existential instantiation and truth-functionally valid transformation rules. In this way, the host of explications of ordinary concepts that make up modern symbolic logic served to set a longstanding problem on a sure footing. Note, however, what kind of problem it was. The horse's head argument represents a compelling counter-example to the Aristotelian syllogistic system precisely because we recognize it, *regardless of Aristotle's system*, as a good argument, *whatever* 'good' amounts to in this context. As interesting and satisfying as it may be to resolve puzzles of this sort, the value of the formal system would be

³⁰ DeMorgan originally posed this challenge to Aristotelian syllogism in the form dealing with the heads of men. The tradition has taken it up as an argument about horse's heads; it's unclear when the DeMorgan's formulation was changed, though the horse's head argument is universally attributed to him.

³¹ Actually, my experience as an undergraduate at a Catholic university taught me that there's still a thriving niche industry in Aristotelian logic apologetics among devout Thomists.

rather limited, were it relegated to formalizing arguments (like the horse's head) we already correctly believe to be valid. Beyond simple-seeming arguments of this sort, on which we feel we have a presystematic grasp and with which any viable formal system would need to agree and would need to clarify, the system's enduring contribution in power it gives us to generate proofs for arguments which are much more complex and about which we don't have the kind of intuitive clarity that we do in the case of the horse's head argument. In this way, the system doesn't just serve to justify previously held beliefs and assumptions, it serves as a systematic tool for the development and acquisition of new knowledge. The entire development of the foundational program and subsequent modern mathematics is a colossal example of this.

Spelling out this development requires retelling the story of the search for a foundation for mathematics as it proceeded in collusion with the formulation of variously ordered systems of logic. In summary, following upon the hopeful notion that all mathematics could be founded on an axiomatization of set theory, a series of proofs using the newly explicated concepts of symbolic deductive systems indicated that such a foundation could not be fundamental. By 1923, Skolem began suggesting that an axiomatization of set theory could not be foundational, because the Löwenheim-Skolem Theorem entails that even within the highly restricted system of first-order logic, any axiomatization of set theory (or the real numbers) would have both a countable and a non-countable model and, thus, fail to be categorical (Moore 1988). What followed over the next decades was a delicate working-out of the relationship between set theory and logic as well as of various varieties of

explicated concepts within logic, effectively adjudicating among competing approaches to logical syntax and semantics as well as questions involving first-order versus second-order and other higher-order systems of formal logic. These developments did not represent, like the horse's head, resolutions of puzzles and problems that had remained outstanding for ages, rather they represent approaches to a familiar field of research that were not previously available, so much so, indeed, that the very questions being posed (to a moderate degree of approximation) weren't even articulable beforehand. It bears noting, in this respect that, after all, the most famous mathematical proofs of the 20th century, Gödel's incompleteness theorems, indicate claims about the power and limitations of systems of formal logic of various types.

In the past two chapters, I have developed a pair of interrelated explications and offered a series of criticisms and suggestions about how these more precise concepts can open new pathways forward in the development of research in cognitive psychology. In this sense, I have tried to offer a radically humble and localized example of what the explications of formal logic systems did for mathematics and its subfields. Still, though, I think there remain other uses for these explications, outside cognitive psychology. One such use is within more traditionally philosophical discussions of analogy. Another falls within the domain of information technology. These further uses of my explications become evident once we start to look at meta-level characteristics of the kind of structures that analogical similarities instantiate. I will detail these applications in the remaining sections of this chapter.

If we recall my clarification of systematicity in Chapter 4³², the lesson there was that qualities of complexes' structures differentiate empirically, not syntactically or formally. Grasping this leads me to the notion that the things we should be looking for in terms of differentiating features between various analogies, especially when we're concerned with relative evidentiary evaluation, are shared structures that exhibit empirically entrenched characteristics. This idea recalls the epistemic story of Goodman's 'grue' paradox (1983). One way of describing that paradox is that it indicates the limitations on purely formal characterizations of inductive confirmation. Neither the set of so far observed emeralds nor any other collection of evidence can "tell you" how to observe and collectively describe it. They're all green, but they also share countless other descriptions not apposite for use in a generalizing hypothesis about emeralds. It is only with continued, successful generalizations using 'green' that such concepts become entrenched in our use and become epistemically viable foci for observation of and theorizing over empirical phenomena.

The difficulty, literally, *of seeing* complex shared structures my explication of analogy makes apparent relates metaphorically to the comparison I've been developing throughout this section. One of the things achievable, for example, regarding the horse's head argument, only after first-order logic was formulated was that the relations were "seeable" by a formalized argument structure, and the relation '___ is the head of ...' is what's doing the logical work in the argument. Aristotelian syllogistic logic can "see" only one relation: '___ is/are...' For this reason,

³² Pp. 167ff

the system will fail to elucidate any argument that does not turn on a subject (or class of them) and its possession (or failure to possess) a predicated quality. Since what's doing the logical work to get the conclusion out of the single premise in the horse's head argument is a set of interrelationships of predicate extensions, both monadic ('horse', 'animal') and dyadic ('is the head of'), the logical structure of the inference fails to be represented in the formal system. It is often noted, for example by DeMorgan (1847) as well as by Russell (1961), that the lesson of the horse's head argument is that, *pace* Aristotle, not all logical arguments are syllogistic, since this argument has only one premise. That may be an indication that something is wrong with Aristotle's system, given that it insists logical arguments must be syllogistic and this argument is not, but the deeper reason the argument presents a challenge to his system is that the relationships between the three predicates that underwrite the inference are not represented symbolically in his system. Any schema of the form 'All As are Bs' implies a set of deep structural relationships. The conclusion of the horse's head argument is one such relationship. However, since the schema fails to syntactically separate the unadorned items being predicated of (the bound variables) from the predicates themselves ('__ is a horse', '__ is an animal'), it provides no syntax for representing these implied relationships. It is in this sense that I mean the system "can't see" the logical structure at work in the argument.

To elaborate these last two ideas, on entrenchment and on relative observability, in the next sections I'll draw on, first, my concept of sameness in a respect and, then, my concept of a structure-predicate. My suggestion is that the first of these notions can elucidate a traditional philosophical preoccupation with

analogies – the validity of analogical arguments – and the second can be used, within a context of digital computation to organize and evaluate vast stores of information in new ways.

II. The Past Is the Beginning and the Middle

A. The Isomorphic Structure in the Age of Universal Availability

One of my repeated complaints about the main line of research on analogical cognition within empirical psychology has been the failure within that field to recognize and incorporate awareness of the universal presence of analogical similarity between compared items. One can choose at will any two things and there will be analogies between them, just not, perhaps, the right kind, the obvious kind, or the kind meant by the people asking the questions. Psychologists seem to view analogies a little like the State Department views terrorism – they know which instances they have in mind, and you just need not bother trying to apply a definition in a consistent way.

Since universal applicability is entailed by my account (and any other explicit account that recognizes shared features as sufficient), my efforts in this project might appear to have an exclusively negative thrust. One might be reasonably tempted to read my whole project as undermining any normative force a judgment or claim of analogy might be supposed to have had, since, on my account, a basic claim of an analogical similarity doesn't narrow down the field at all. To answer such a reading, comparison to the 'grue' paradox and its lessons is again apposite. A kind of parallel argument holds for induction itself, and it remains below the surface of

the ‘grue’ argument as an extreme case. The ‘grue’ argument compares two sets of observations: the set observing each emerald to be green and the set observing each emerald to be grue. However, we might also compare another set of observations from a consistent evidence source: the observation of each emeruby to be gred³³. The presence of emerubies highlights the fact that strange and unfamiliar predicates can denote not just properties, like color, but the members of any type of compliance-class. Thus, it is possible to end up drawing all manner of disparate objects, like all future-past-emeralds and future-future-rubies, together as a type and truthfully predicate some characteristic of each of them. Since this is the case, the process of classification and joint predication upon which the universal generalizations of induction are based is itself subject to the kind of radical overdetermination characteristic of analogies as I have explicated them. Just as, with enough ingenuity and weirdness, it is possible to describe a shared abstract structure between any pair of items, it is also possible to find common features between any set of items such as to ground an inductive generalization. In the case of induction, the motivation was never to suggest that, as a consequence of this overdetermination, no such generalizations ever carry normative force. Rather, it was to highlight the role that past and continuing empirical engagement play in developing and strengthening this force, by degrees. The radical overdetermination of shared structures warrants a similar, though importantly different response.

With induction, the role played by the empirical is directly semantic. By this, I mean that past inductive practice, as a primary informant to our ongoing

³³ Anything that is either an emerald and observed before time *t* or a ruby and not so observed is green and observed before time *t* or red and not so observed.

engagement with the world, constrains the range of reasonable semantic interpretations of abstract predicate schemata within a formalized characterization of inductive inference. Any basic inductive inference of the sort I have been discussing would be schematized as follows:

$Ea \bullet Ga \bullet Eb \bullet Gb \bullet Ec \bullet Gc \dots$

Confirms: $\forall x (Ex \rightarrow Gx)$

Past and continuing deductive practice constrains the range of available interpretations for the predicate schemata 'E' and 'G'. Though, ranging over the same body of evidence, the premise remains true across various interpretations of the predicate letters ('E' as 'emerald' and 'G' as 'green' versus 'G' as 'grue'), the confirmation is legitimate only under the first interpretation, because past practice has not only not given us any reason to regard 'grue' as a regular natural characteristic, there are no other natural characteristics like it at all.

The overdetermination of analogical similarity should be handled in a way similar but also importantly different from this. In the case of analogies as they function in *arguments*, which has been the major focus of work on analogies within philosophy, the terms to be entrenched are not the predicates appearing in the shared structures so much as the respects of sameness that underwrite this sharing. I will now turn to fill out and explain this idea. After that, I will offer some more speculative remarks about how my theory could ground more novel approaches to organizing and evaluating complex digital data-sets by developing relative entrenchments of various structure-predicates through ongoing investigative use.

II. Analogical Arguments

A. The Isomorphism Approach: Its Attractiveness and Indeterminacy

In the philosophical literature on the logic of analogical arguments, two main approaches have been suggested. The first, which follows the Carnapian tradition of formal explication of induction through a probability function, operates on the basic idea that analogues known to share a property are more likely to share a further property. I call this the ‘shared properties view’. The other approach, about which very little has been said in the literature, was introduced by Julian Weitzenfeld (1984), in his classic paper “Valid Reasoning by Analogy”. Weitzenfeld’s view holds that analogical arguments are valid insofar as there are underlying isomorphic structures between the base and the target domain that legitimate the drawing of the specific conclusions in question. The problem with the first of these two views is that the sharing of certain properties is irrelevant to the sharing of certain others, a condition which does not lessen in severity with the accumulation of more and more properties. So, in many cases, the fact that two analogues are alike in a stated way may be entirely irrelevant to whether the target object shares a further projected property in common with the base analogue. The second, the isomorphism approach, does not obviously have the problem of the former, because, when analogues are structurally isomorphic, certain features or properties are structurally relevant to others. The structural isomorphism approach, thus, improves on the shared properties view, but it has its problems, which can be lessened by an emendation of what I term ‘respectivism’.

One issue central to the isomorphism approach is that, without some qualification, it makes analogies obsolete: in order to apply it, you must know that the target analogue is isomorphic to the base analogue; but in order to know that, you must know the relevant structure of the target analogue, and, if you know that, then you can get to the conclusion directly. So, in order for the structural isomorphism approach to be illuminating, it must be glossed so that the structural features of the target need not be known in advance. Respectivism, of the sort that I recommend, can help do this.

The other main issue that must be dealt with from the structural isomorphism standpoint is that analogical arguments usually are not stated in the terminology of structural isomorphism; they are stated in ordinary language. And, for any analogical comparison between two things (statements of the form 'a is like b', or 'a and b are both P's'), there will be a countless number of isomorphic structures shared between the base and the target that are consistent with and (partially) described by the analogy or comparison, as stated. Some of these will structurally legitimate the projection of the predicted property onto the target, and some will not. Most that do not are bound to be weird, but be isomorphic nonetheless. The problem, then, for the isomorphism approach is that it doesn't have a way of tagging the language of the argument to the useful isomorphic structures only. Like the first of the two central problems with the structural isomorphism account, my approach, using respects of sameness, helps patch this gap.

To see how this could work, I'll discuss a famous example of the successful use of analogical argumentation in actual scientific practice. In his *Astronomia Nova* (1609), Kepler wanted to defend the Copernican view of cosmology, but struggled with the fact that it seemed to admit action-at-a-distance. If the sun were the center of the universe and the cause of the motion of the earth around it, it would seem to have to act on the earth without any observable physical interaction. In order to make the idea that the sun causes this motion by a "motive power" plausible, Kepler drew an analogy between sunlight and motive power. Unobjectionably, the Sun is the cause of daylight on Earth, yet there is no observation of light in the spaces between the Sun and Earth. While the analogy leaves all sorts of aspects of planetary motion unexplained, it does effectively move what initially seems implausible into the realm of plausibility and leads to further routes for inquiry and the formulation of hypotheses. Taking the structural isomorphism approach, Kepler's argument can be analyzed as involving two systems, I'll call them 'Sunlight₁' and 'Motion₁'. Here's a first attempt at the right sort of structural isomorphism; I'll call it

'I_{SM1}':

Sunlight₁:

Motion₁:

U _S : {Sun, Earth, light}	U _M : {Sun, Earth, motive power}
Comes From: {⟨light, Sun⟩}	Comes From: {⟨motive power, Sun⟩}
Effects: {⟨light, Earth⟩}	Effects: {⟨motive power, Earth⟩}
Unobservable between: {⟨light, Sun, Earth⟩}	Unobservable between: {⟨motive power, Sun, Earth⟩}

Something like the above structural isomorphism between the systems, Sunlight₁ and Motion₁, is supposed to obtain so as to provide evidence that motive power and its source in the sun is a plausible explanation of how the sun could cause the motion of the Earth without the sun “acting-at-a-distance” in any strange way. The elements of Sunlight₁ can be mapped 1-to-1 onto the elements of Motion₁ so that all of the relevant relations between the sun, earth, and light are preserved between the sun, earth, and motive power; and, since the third relation of the set ‘unobservable between’ isn’t a problem for the first system, it shouldn’t be a problem for the second system.

Nevertheless, if we look at how the analogy is stated, what commonalities are identified between the systems and the ways in which they are identified to be the same, it becomes clear that the foregoing identification of the isomorphism between the systems leaps many epistemic hurdles. That is, the isomorphism identified is consistent with the stated commonalities between the base and the target systems, but it is not the only such isomorphism, and nothing about those commonalities circumscribes just that one and no others. Here’s Kepler’s argument:

But lest I appear to philosophize with excessive insolence, I shall propose to the reader the clearly authentic example of light, since it also makes its nest in the sun, thence to break forth into the whole world as a companion to this motive power. Who, I ask, will say that light is something material? Nevertheless, it carves out its operations with respect to place, suffers alteration, is reflected and refracted, and assumes quantities so as to be dense or rare, and to be capable of being taken as a surface wherever it falls upon something illuminable. Now just as it is said in optics, that light does not exist in the intermediate space between the source and the illuminable, this is equally true of the motive power. (Kepler 1609: 383)

Now, looking at Kepler's own statement of the analogical argument, we see some differences from how the proponents of the structural isomorphic view would interpret the argument. One way of spelling out the argument as stated is that it tries to establish that motive power is "authentic", given that light is authentic and light (l) and motive power (m) are similar in various ways. In particular, light "does not exist in the intermediate space between the source and the illuminable." Some of the ways delineated in the passage are as follows: l and m are both caused by the sun; l and m are both non-material; l and m both appear at specific places, change, and have observable properties; l and m both do not exist in the space between their sources and the objects upon which they act. These similarities, along with the assumed fact the light is "authentic", constitute a ready paraphrase of the evidence given in Kepler's argument.

Important questions arise out of the comparison of the flat paraphrase and I_{SM1} above, which must be identified in order to apply the structural isomorphism analysis of the argument. First of all, Kepler's argument compares light and motive power. I_{SM1} does not specifically compare these items, but rather compares two structures, of which light is a member of one and motive power is a member of the other. That, in itself, may not be too serious, if the resources for generating the pair of structures are either in or implied by the premises as stated. So, the question, then, is – can the structural isomorphism used in the analysis of the argument (I_{SM1}) be gleaned from the information stated in the argument?

Not quite, since both structures in I_{SM1} contain just three things, one of which (earth) is not mentioned explicitly in the argument, and both feature the relation

‘unobservable between’, which is not in Kepler’s argument. Kepler talks rather about *l* and *m* *not existing* between the sun and the things they affect, rather than being merely unobservable. If I_{SM1} , in particular, though, cannot be gleaned from Kepler’s argument, is it possible that some other relevantly isomorphic pair of structures can, such that the argument goes through in virtue of that pair? I think so. The problem with this, however, is that the argument determines *far too many* such structures, many of which will be useless to appeal to for support for the argument. Here’s I_{SM2} , based directly on the evidence stated in Kepler’s argument:

Sunlight₂:

Motion₂:

$U_{S2}: \{\text{Sun, light, } i_1, i_2, i_3, \dots\}^{34}$	$U_{M2}: \{\text{Sun, motive power, } m_1, m_2, m_3, \dots\}$
Comes From: $\{\langle \text{light, Sun} \rangle\}$	Comes From: $\{\langle \text{motive power, Sun} \rangle\}$
Affects: $\{\langle \text{light, } i_1 \rangle, \langle \text{light, } i_2 \rangle, \langle \text{light, } i_3 \rangle, \dots\}$	Affects: $\{\langle \text{motive power, } m_1 \rangle, \langle \text{motive power, } m_2 \rangle, \langle \text{motive power, } m_3 \rangle, \dots\}$
Nonexistent between: $\{\langle \text{light, Sun, } i_1 \rangle, \langle \text{light, Sun, } i_2 \rangle, \langle \text{light, Sun, } i_3 \rangle, \dots\}$	Nonexistent between: $\{\langle \text{motive power, Sun, } m_1 \rangle, \langle \text{motive power, Sun, } m_2 \rangle, \langle \text{motive power, Sun, } m_3 \rangle\}$
Authentic: $\{\text{Sun, light, } i_1, i_2, i_3, \dots\}$	Authentic: $\{\text{Sun, motive power, } i_1, i_2, i_3, \dots\}$

More of what Kepler adduces as evidence could be added to this, but this pair of structures gets across substantially what he’s trying to show in his argument. The several similarities he points out suggest a structural isomorphism between certain physical systems, one of which contains light, the other of which contains motive power; were such a structural similarity to actually obtain, then whatever is true of

³⁴ Here, ‘*i_n*’ stands for ‘illuminable thing’, drawn from Kepler’s statement that light is “capable of being taken as a surface wherever it falls upon something illuminable”. The analogue in U_{M2} (‘*m_n*’) stands for something like ‘movable thing’, understood as just those things which are capable of being moved by motive power.

light in the one structure, will be true of its analogue in the other structure – in this case, motive power. One of the predicates denoting light in the first structure is ‘authentic’, so, by the isomorphism relation, motive power is also shown to be “authentic”.

However, at this point, when the structure pair attempts to remain more faithful to the word of the argument, isomorphism runs aground. This immediately becomes evident if we look at the second relation ‘affects’, in each of the structures. ‘Affects’ in Sunlight₂ does not map 1-to-1 onto ‘affects’ in motion₂, because there are important differences between illuminables and movables: *all* movables are acted on by motive power – even those in outer space – but some illuminables only in fact reflect light potentially. Cave dwellers can hide from light; but from gravity – the “motive power” of the sun, there’s no place to hide. Since this is the case, there will be members of $m_1...m_n$ in the second position of ordered pairs in the extension of ‘affects’ in Motion₂ that have no correlating $i_1...i_j$ in the second position of ordered pairs in the extension of ‘affects’ in Sunlight₂. Or, put more simply, some things will be affected by motive power which do not have unique correlates that are in fact affected by light. This undermines structural isomorphism.

So, let me recap my argument up to this point. One pair of structures was considered, I_{SM1} , which fits roughly with a colloquial description of Kepler’s argument. However, once we look at the argument itself, it becomes clear that although I_{SM1} is consistent with it, Kepler’s actual argument is more general, and involves not just Earth, but all things that are affected by the Sun’s light and motive power. When we change the structure pair to I_{SM2} to accommodate this generality,

the structures are no longer actually isomorphic to each other, although the pair of structures themselves faithfully reflect Kepler's original argument.

Suppose we try to find some sort of rough middle way, say, including just the planets as the illuminables and movables, since planets aren't often in the dark. This would help prevent the difficulty of ending up with non-isomorphic structures, but it is still open to another difficulty falling out of the general problem of mapping arguments onto structures. It is possible to generate pairs of isomorphic structures that respect the evidence given in Kepler's argument but which do not support the conclusion that motive power is "authentic". I_{SM3} is one such pair:

Sunlight₃:

Motion₃:

U _{S3} : {Sun, light, Mercury, Venus, Earth, Mars, Jupiter, Saturn} ³⁵	U _{M3} : {Sun, motive power, Mercury, Venus, Earth, Mars, Jupiter, Saturn}
Comes From: {⟨light, Sun⟩}	Comes From: {⟨motive power, Sun⟩}
Affects: {⟨light, Mercury⟩, ⟨light, Venus⟩, ⟨light, Earth⟩, ⟨light, Mars⟩, ⟨light, Mars⟩, ⟨light, Jupiter⟩, ⟨light, Saturn⟩}	Affects: {⟨motive power, Mercury⟩, ⟨motive power, Venus⟩, ⟨motive power, Earth⟩, ⟨motive power, Mars⟩, ⟨motive power, Mars⟩, ⟨motive power, Jupiter⟩, ⟨motive power, Saturn⟩}
Nonexistent between: {⟨light, Sun, Mercury⟩, ⟨light, Sun, Venus⟩, ⟨light, Sun, Earth⟩, ⟨light, Sun, Mars⟩, ⟨light, Sun, Venus⟩, ⟨light, Sun, Jupiter⟩, ⟨light, Sun, Saturn⟩}	Nonexistent between: {⟨motive power, Sun, Mercury⟩, ⟨motive power, Sun, Venus⟩, ⟨motive power, Sun, Earth⟩, ⟨motive power, Sun, Mars⟩, ⟨motive power, Sun, Venus⟩, ⟨motive power, Sun, Jupiter⟩, ⟨motive power, Sun, Saturn⟩}
Authentic: {Sun, light, Mercury, Venus, Earth, Mars, Jupiter, Saturn}	Referred to by Kepler: {Sun, motive power, Mercury, Venus, Earth, Mars, Jupiter, Saturn}
Material object: {Sun, Mercury, Venus, Earth, Mars, Jupiter, Saturn}	Authentic: {Sun, Mercury, Venus, Earth, Mars, Jupiter, Saturn}

This pair of structures, unlike those in I_{SM2}, is genuinely isomorphic, and it can be reasonably gleaned from the evidence offered in Kepler's discussion of light and motive power. It does not, however, substantiate the claim that motive power is "authentic". This represents what I take to be the second main problem with the structural isomorphism approach to analogical arguments: analogical arguments determine many, many isomorphic structures between systems; so isomorphic structures represented by similarity evidence cannot be sufficient for successful

³⁵ Kepler knew of only the first six planets.

analogical arguments. And, since there are bound to be such structures for any pair of analogues, noting that they are necessary fails to provide an explanation that narrows things down at all. A pair of structures that *would* legitimate the inference (I_{SM4}) would be identical to I_{SM3} except the fourth predicate of $Motion_4$ would be 'authentic' instead of 'referred to by Kepler', and there would be no fifth predicate. I will now argue that entrenched respects of sameness can be used to single out this (or the other useful pairs of isomorphic structures) from others such pairs that are implied by the evidence but irrelevant to the argument (pairs like I_{SM3}).

B. Respectivism Added

In the first section, I argued that the isomorphism strategy for analyzing analogical arguments has significant attractions over the "shared properties" approach, but that, due to underdetermination of *specific* cases of isomorphism by analogical arguments, as stated, a structural isomorphism cannot be sufficient for satisfactory explication; and due to overdetermination of isomorphism in general between two systems containing the analogues in question, the necessity of isomorphism is only limitedly illuminating. So, it seems that something must be added to the isomorphism account if it is to be useful. Moreover, an extreme consequence of this failing of the isomorphism approach is that it seems to make the analogy itself obsolete, since its application seems to require knowledge of the relevant structures for the target analogue, and having such knowledge would be sufficient on its own to draw or reject the projected conclusions, without consideration of the analogical relation at all.

If we add respectivism to the analysis of analogical arguments by structural isomorphism, these problems can largely be solved. To show how this works, let me first review what I take respectivism to be.

Respectivism relies on my explication of the sameness relation, which is given entirely in terms of the denotation relation between predicates and objects they denote. Two things are the same with respect to r *just in case* there is some r -predicate that denotes the two things in common. So, for example, light and motive power are the same with respect to source just insofar as the source-predicate (say, 'comes from the sun') denotes them in common. This account can also be generalized to include sameness relations between ordered n -tuples of objects: two ordered n -tuples of objects are the same with respect to r *just in case* there is some n -place r -predicate that denotes the two n -tuples in common. Given this, the example above can be altered to exactly fit the relation in the structure I_{SM4} : $\langle \text{light, sun} \rangle$ and $\langle \text{motive power, sun} \rangle$ are the same with respect to source designation, insofar as the relation 'comes from' denotes each pair and it's itself a source designation-predicate.

Using this explication, it is possible to view structures that exemplify isomorphism as sets of sameness relations. And these are actual sameness relations, with some empirical significance and subject to the ordinary sort of varied degrees of entrenchment that other semantic items are by inculcation into empirical practice. This relative entrenchment cannot fully undermine the indeterminacies of the structural isomorphism approach, but it does go some considerable distance

toward doing so. But this is all getting rather abstract; let's return to the example to fill out the claim.

Here's I_{SM4} :

Sunlight₄:

Motion₄:

U_{S4} : {Sun, light, Mercury, Venus, Earth, Mars, Saturn, Jupiter}	U_{M4} : {Sun, motive power, Mercury, Venus, Earth, Mars, Saturn, Jupiter}
Comes From: {⟨light, Sun⟩}	Comes From: {⟨motive power, Sun⟩}
Affects: {⟨light, Mercury⟩, ⟨light, Venus⟩, ⟨light, Earth⟩, ⟨light, Mars⟩, ⟨light, Mars⟩, ⟨light, Saturn⟩, ⟨light, Jupiter⟩}	Affects: {⟨motive power, Mercury⟩, ⟨motive power, Venus⟩, ⟨motive power, Earth⟩, ⟨motive power, Mars⟩, ⟨motive power, Mars⟩, ⟨motive power, Saturn⟩, ⟨motive power, Jupiter⟩}
Nonexistent between: {⟨light, Sun, Mercury⟩, ⟨light, Sun, Venus⟩, ⟨light, Sun, Earth⟩, ⟨light, Sun, Mars⟩, ⟨light, Sun, Venus⟩, ⟨light, Sun, Saturn⟩, ⟨light, Sun, Jupiter⟩}	Nonexistent between: {⟨motive power, Sun, Mercury⟩, ⟨motive power, Sun, Venus⟩, ⟨motive power, Sun, Earth⟩, ⟨motive power, Sun, Mars⟩, ⟨motive power, Sun, Venus⟩, ⟨motive power, Sun, Saturn⟩, ⟨motive power, Sun, Jupiter⟩}
Authentic: {Sun, light, Mercury, Venus, Earth, Mars, Saturn, Jupiter}	Authentic: {Sun, motive power, Mercury, Venus, Earth, Mars, Saturn, Jupiter}

Given the structural isomorphism between Sunlight₄ and Motion₄, all sorts of sameness relations obtain between light and motive power as well as between ordered n -tuples of objects which have light and motive power as members. And none of the respects in which these objects and n -tuples are the same is particularly weird. For example, though the “respects” formulation is linguistically unusual, there's nothing empirically odd about inquiring after or caring about whether two things are the same with respect to designated origin, or cause; or whether things

are the same with respect to existence, or, in Kepler's terminology, "authenticity". These sorts of similarities exemplify relevant points of empirical concern and are good indications of relevant structural similarities. Knowing that two things are alike in these ways is certainly not logically decisive, but analogical arguments are inductive, so that shouldn't be a mark against its relevance. What knowledge of these sorts of similarities does, especially when they are combined with other similarities so as to comprise a pair of isomorphic structures, is to legitimate probabilistic inferences that rely on structural similarities. This legitimacy comes from the accumulation of past successes, just as in inductive inference generally.

At this point, all I have discussed is how respectivism can work in tandem with the structural isomorphism view. However, as I have suggested, it can also help that view overcome its problems. So the case I have discussed is one where an appropriate pair of structures is identified from the stated argument. But the central problem I noted with the view is that there are too many such structure-pairs, the vast majority of which are *not* appropriate. Respectivism, with its requirement of relative entrenchment, can help differentiate the appropriate from the inappropriate structures. To show this, let us again consider the inappropriate structural pair I_{SM3} .

As we saw above, I_{SM3} , culled from Kepler's argument, not only fails to legitimate the conclusion that motive power is "authentic", but it also confirms that motive power is non-"authentic". The two aspects of the isomorphic pair of structures responsive for these outcomes are the mapping of 'authentic' and 'material object' in $Sunlight_3$ onto 'referred to by Kepler' and 'authentic' in $Motion_3$,

respectively. However, whatever sameness relations these two mappings represent, they are bound to be sameness in *weird* respects, unlike sameness with respect to designated origin or with respect to possible existence. Trying to think up respects of sameness to fill these spots shows just how empirically circumstantial these similarities are. For example, unlike in the case of I_{SM4} and ‘comes from’ or ‘authentic’, in the case in question, we *don’t* have a normal predicate denoting the two items in common. So, the predicate we use in the explication will have to be disjunctive. The predicate ‘authentic or referred to by Kepler’ denotes both light and motive power. But what sort of predicate is *that*? I suppose we could call it an “existence-and-reference-report-predicate”, since it’s about both existence and the reporting of someone’s referring to something. But ascriptions of sameness with respect to “existence-and-reference-report” do not appear to identify empirically interesting similarities. As far as I’m aware, in the history of empirical thought, no such similarities have ever represented useful structural relations or empirical patterns, and so this respect in which things can be the same is not a respect that we can identify as entrenched in analogical practice. This is not to say that such a respect *could not* become thus entrenched; it is only to say that, by our current best empirical lights, similarity with respect to existence-and-reference-report has appeared to be a coincidental respect of similarity and has not so far grounded successful projections onto future cases. The only evidence I have to support this claim is that such a respect of sameness appears to me to be *weird*; I’ve never heard of any type of empirical generalization successfully grounding inferences to future cases based on the conjunction of ascription of existence to one (type of) thing with

the fact that someone referred to some (other) thing. Mere existence just doesn't provide detailed empirical evidence, and the fact the someone referred to something doesn't appear to be conditioned (even statistically) by any currently understood scientific laws; moreover, the conjunction of two these two types of facts appears to be just that much more circumstantial.

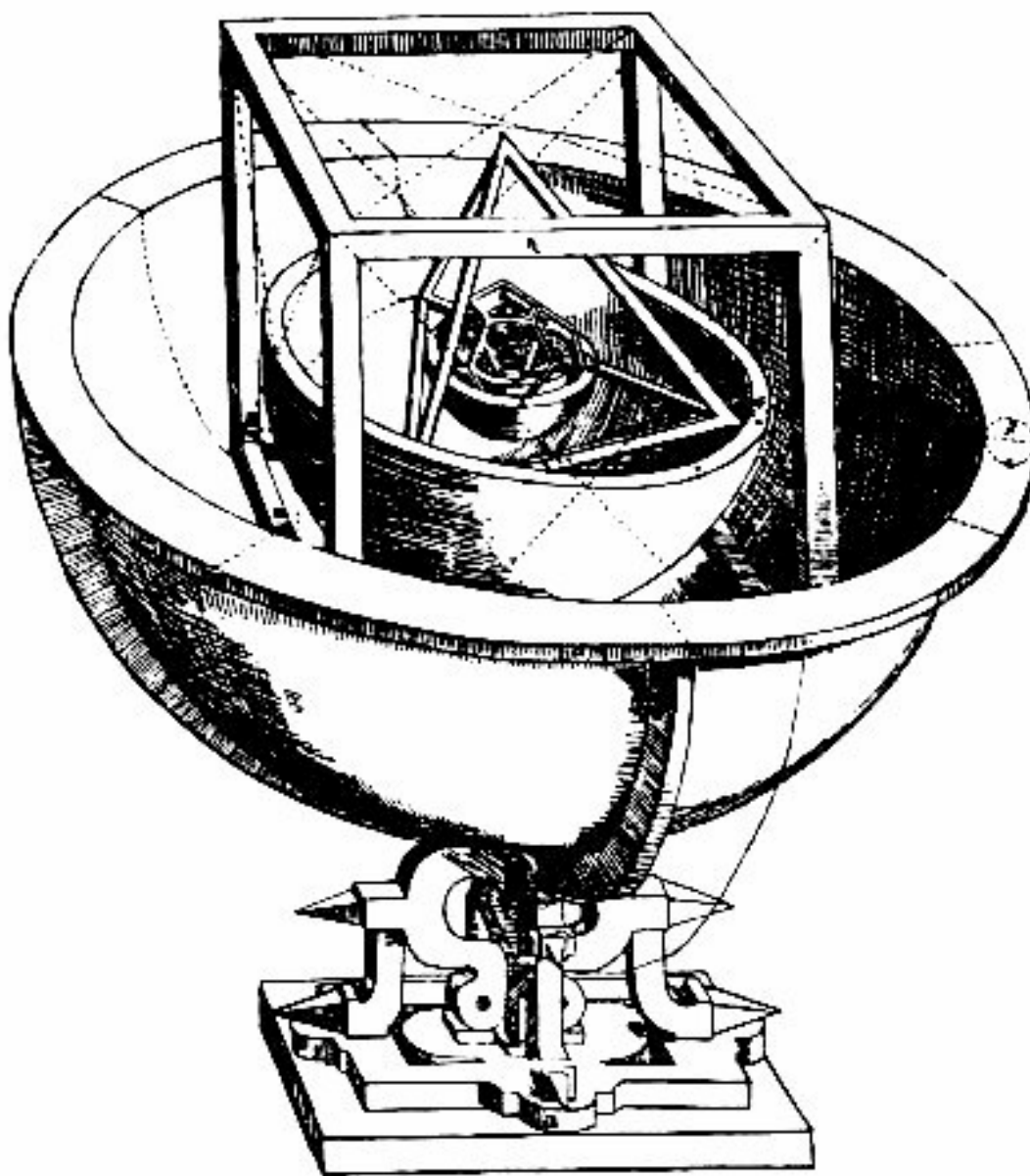
The same kind of treatment can be applied to the other problematic aspect of I_{SM3} . 'Material object' does not denote light, and 'authentic' does not denote motive power, so the predicate 'not both material and authentic' denotes both light and motive power in common. 'Not both material and authentic' is not an ordinary sort of predicate either; the best description of it that I can think of is a "metaphysical status/existence conjunction denial-predicate". This is not a respect in which the sameness of two things has significant empirical consequences; it is a gerrymandered, instead of an entrenched, respect.

So, having discussed both the example of the "correct" isomorphism to focus on and the incorrect pair that can be drawn from Kepler's argument, I have shown that my respectivism has a way to focus the questions relevant for sorting out the useful from the useless cases of isomorphism between any two structures containing the analogues. An isomorphic structure is useful *just in case* the respects of the sameness relations that obtain between the analogues are relatively entrenched. "Entrenchment" is an informal notion, which appeals directly to common usage and sense. Some attempts have been made to rigorously specify the difference between entrenched predicates (or ones that are useful in inductive contexts) and non-entrenched predicates. No such attempts have ever outstripped

the need for the application of ordinary sense in distinguishing between useful and weird predicates. Furthermore, the program that these attempts represent can be seen to be of a piece with the structural isomorphism approach without my tempering addition of respectivism. I will return to these issues in my summary.

C. Application to a Case of Bad Analogical Reasoning

Just as, without the respectivism constraint added, the structural isomorphism approach lacks resources for differentiating appropriate from inappropriate isomorphism, and so fails to explain the legitimate drawing of a conclusion in cases of analogical argumentation that actually work (like Kepler's motive force argument), it also legitimates the conclusion in cases that don't work. Kepler was an avid analogy-maker, and not all of his analogical arguments were good ones. In *Mysterium Cosmographicum* (1596), Kepler made many analogical arguments about the orbits of the planets by comparison to spheres and the five Platonic solids (solids made of perfectly symmetrical non-co-planer point sets). The two central arguments Kepler made using this comparison, were that the number of planets is six, because the number of perfect solids is five; and the distances between the planets are what they are because of the ways in which the solids can both contain and be circumscribed by spheres. These arguments correspond to a cosmographic model Kepler designed such that the orbit of each of the planets falls along a sphere that is circumscribed around one of the regular solids, with the innermost sphere being inscribed in the innermost solid. Here is Kepler's drawing of the model:



With the help of a little delicate handling, a complex pair of structures can be determined to exemplify the relevant claims corresponding to this model. But before discussing that, I'll provide a little background for his approach.

Kepler's approach draws on Pythagorean, Platonic, Aristotelian and Christian theological ideas. For him, the sphere, being a "perfect" form, represents the triune God. The center point is the father, the points along its surface are the son, and the

consistent equidistance between the center point and the surface points represent the spirit. Given this view of God and his relation to the sphere, Kepler suggests that, in order to have numerable things in the universe there must be differentiation, and the geometrically differentiated counterpart to the curve (instantiated in the sphere) is the line. Thus, he believed that God, in being perfect like the curved sphere must use a non-curved object to bring differentiation—the various astral bodies—into existence. But this differentiation approximates God’s perfection and so must fit exactly within the spheres of the universe; the solids that do this are the five perfect, or “Platonic” solids: cube, tetrahedron, dodecahedron, icosahedron, and octahedron. These, as well as other considerations pertaining to the particularities of the solids themselves, serve as Kepler’s motivation for the model he proposes.

The whole of *Mysterium Cosmographicum* is what Darwin sometimes called his *Origin of Species*: one long argument, so there is no concise statement of the argument in the text. However, here is a passage that gathers together much of his reasoning in the general formulation of his geometric model:

For just as God is the model and rule for living creatures, so the sphere is for solids. Now the sphere has the following properties: 1. It is extremely simple, because it is enclosed by a single boundary, namely itself. 2. All its points are at a precisely equal distance from the center. Therefore, among bodies the regular solids approach most closely to the perfection of the sphere. Their definition is that they have: 1. All their edges, 2. their faces, and 3. their vertices respectively equal both in kind and in size, which is a sign of simplicity. From the adoption of this definition it follows automatically that 4. the centers of all the faces are equally distant from the midpoint, 5. That if they are inscribed in a globe all their vertices touch its surface, 6. That they are fixed within it, 7. That they touch the inscribed globe at all centers of their faces, 8. That consequently the inscribed globe is fixed and immobile, 9. And that it has the same center as the solid. These properties yield another resemblance to the sphere, which results from the equality of the distances between the faces. (Kepler 1596: 101)

Here is a pair of structures that captures some of the relevant aspects of his model and its relation to the planets' orbit: I_{SP1} :

Solids₁

Planets₁

U _{s1} : {center point, s(phere)1, s2, s3, s4, s5, s6}	U _{p1} : {Sun's center, Mercury's orbit, Venus' orbit, Earth's orbit, Mars' orbit, Jupiter's orbit, Saturn's orbit.}
All points on ___ are closer to than ---- : {⟨s1, center point, s2⟩, ⟨s2, center point, s3⟩, ⟨s3, center point, s4⟩, ⟨s4, center point, s5⟩, ⟨s5, center point, s6⟩, [and all other permutations following from the transitivity of 'closer than'.]}	All points on ___ are closer to than ---- : {⟨Mercury's orbit, Sun's center, Venus' orbit⟩, ⟨Venus' orbit, Sun's center, Earth's orbit⟩, ⟨Earth's orbit, Sun's center, Mars' orbit⟩, ⟨Mars' orbit, Sun's center, Jupiter's orbit⟩, ⟨Jupiter's orbit, Sun's center, Saturn's orbit⟩, [and all other permutations following from the transitivity of 'closer than'.]}
___ is circumscribed by an octahedron that inscribes : {⟨s1, s2⟩}	___ is contained _m by ... ³⁶ : {⟨Mercury's orbit, Venus' orbit⟩}
___ is circumscribed by an icosahedron that inscribes : {⟨s2, s3⟩}	___ is contained _v by ... : {⟨Venus' orbit, Earth's orbit⟩}
___ is circumscribed by a dodecahedron that inscribes : {⟨s3, s4⟩}	___ is contained _e by ... : {⟨Earth's, Mars' orbit⟩}
___ is circumscribed by a tetrahedron that inscribes ... : {⟨s4, s5⟩}	___ is contained _r by ... : {⟨Mars' orbit, Jupiter's orbit⟩}
___ is circumscribed by a hexahedron that inscribes ... : {⟨s5, s6⟩}	___ is contained _j by ... : {⟨Jupiter's orbit, Saturn's orbit⟩}

³⁶ To preserve isomorphism, a specific relation must be formulated for each of the orbits to its neighbors. If the relation were a general notion of containment then, for example, Jupiter's orbit would contain Earth's orbit. The corresponding pair in the extension of the general containment relation ((Earth's orbit, Jupiter's orbit)) would have no counterpart in the Solids₁ structure: there is no simple relation in that structure in whose extension falls ⟨s3, s5⟩. Adding a complex relation that would have this value is certainly possible, but it would, of course, be distinct from the other relations between the spheres, exactly what having a general notion of containment in the Planet₁ structure is designed to avoid.

This pair of structures preserves the relations of ordering between the spheres, on the one hand, and the planets' orbit, on the other. The first relation in each structure establishes the order of the spheres and the orbits, respectively; and the other relations can be used to calculate the particular distances between each of the orbits, using a starting measurement of the distance between the center point of the sun and the orbit of Mercury. The relevant ratios are recoverable from the geometric relations contained in Solids₁. So, using the model of analogical reasoning advocated by the isomorphic structure view, many of the great many conclusions drawn by Kepler from this model of the universe could be substantiated. This doesn't speak well for that view, then, since the argument is a bad one.

To be clear, I do not suggest that the analogy Kepler is using is a bad one simply because the predictions he made using it turned out to be wrong. Lots of good inductive arguments, analogical and otherwise, turn out to have false conclusions. One example, in particulars and details, is Kepler's own argument I discuss above, as an example of a good analogical argument – the argument for the plausibility of “motive power”, originating in the sun, as a possible explanation for the motion of the planets. Much more interesting than focusing on the fact that Kepler's model made false predictions about the spacing of the planets, is to note and notice that his predictions were very nearly correct. Given this, we should ask what would be the appropriate response had the predictions had been exactly correct? The worry this question raises strikes to the heart of why the isomorphism view of analogical arguments must be augmented with some focus on embedded

characteristics of practical empirical enquiry. If the predictions had been exactly correct, that alone would not change the status of the argument, because that alone would not entrench the respects in which the planets' orbits and the model's spheres are the same, so the use of the isomorphism *even if it were to deliver veridical data* would be illegitimate. Given the fact that geometrical perfection has nothing, in fact, to do with the placement of bodies in space, the respects of sameness that underlie the isomorphism between Solids₁ and Planets₁, these respects would not be able to gain empirical purchase in further observation and experiment. The results of Kepler, then, had they been even much more approximately correct than they in fact were, could be regarded as nothing more than an elaborate and unlikely coincidence. The relative placement of objects in the natural world may have something to do with and be similar to some sorts of geometric relations, but the relations and properties relied on by Kepler – such as “perfectness” – are, by our best empirical lights, irrelevant to such placement.

This last point, that we criticize Kepler's analogy by *our* best empirical lights, warrants some elaboration. One consequence of this is that Kepler might both accept my theory of analogy and also reject my judgment that his analogical argument is a bad one. This is because my view has it that the relative validity of analogical arguments is always evaluated and judged according to our best background beliefs, up to and including broadly general and theoretical beliefs about how the world works. Since we reject Kepler's analogy on such grounds – grounds he would not accept – he could consistently accept my theory while also use it to judge his own analogy differently from how we judge it. Given this, the

humbleness of my view should be apparent. It provides a heuristic for systematically evaluating arguments that operate by appeal to shared structures, but it does not claim to fill in the point of view from which such evaluations are made; no abstract schema could do this. We will generally come to such tasks from our own points of view, with all of our own background beliefs and outfitted with what we judge to be the most reliable information. However, given the role that background theories play in using the account to evaluate cases of analogical argumentation, the view also provides a framework within which we can adopt alternative points of view for evaluating an analogy. So, for example, we can adopt some approximation of Kepler's point of view when we recognize that the analogy would not have conflicted with his background theories, even though it does conflict with ours. This is not to say, then, that the view entails or advocates historical relativism about theories, or scientific explanations or even about whether an analogical argument is a good or a bad one; rather, it suggests a relativism about when a person would be right to *regard* an analogy as a good one, given her background commitments. Thus, we can explain why Kepler regarded his analogy as a good one while also judging it to be a bad one.

To make the criticism explicit in the terms I have already filled out for the motive power case, one sameness relation that the isomorphic structures instantiate is that the pair $\langle \text{Jupiter's orbit, Saturn's orbit} \rangle$ are the same as the pair $\langle s_5, s_6 \rangle$ with respect to inscription and circumscription *and* containment. Both pairs are denoted by the complex predicate “___ is circumscribed by a hexahedron that inscribes ... or ___ is contained by ...”, which is itself an “inscription and circumscription *and*

containment-relation (predicate).” As far as I’m aware, the respect “inscription and circumscription *and* containment” isn’t something that has any prior empirical use.³⁷ Instances of inscription and circumscription may also, in general, be instances of containment, but the converse is not true; there’s lots of containing going on in nature that has nothing to do with the exact geometric relations of inscription and circumscription. For this reason alone – that the proposed set of isomorphic structures relies on the comparison of items in respects that have no prior empirical significance and so are not at all entrenched in inductive practice, we can disqualify it and the argument it seems to legitimate.

D. Detraction

Some sleight of hand may seem to be afoot. All of the cases where the isomorphism can be dismissed, such as I_{SM3} and I_{SP1} , are cases where a mapping represents a sameness relation that can only be described with the vocabulary of the isomorphism by using a disjunctive predicate. Surely, though some cases of argument-legitimizing isomorphism will be like this, where, for example, a relation in one structure maps onto a *different* relation in the isomorphic structure, so that, in order to generate sameness statements about members of those relations’ extensions, it is necessary to use a disjunctive predicate. Virtually *any* legitimate

³⁷ My not being aware of any prior empirical use of the respect “inscription and circumscription and containment” isn’t certain evidence that this respect has no prior empirical use, given that I might just be poorly informed. Nevertheless, ‘inscription and circumscription and containment’ appears nowhere on the World Wide Web or in any printed or electronic text in the number of major research libraries worldwide that participate in the Google Books Library Project.

mathematical, computer, or scale model of a physical system will be this way³⁸, so it cannot be that such predicates indicate *tout court* the illegitimacy of such isomorphism.

This counterpoint can be answered. There are a couple of possible ways that these sorts of disjunctive predicates could show up; sometimes the respects in which the compared items are the same are entrenched in empirical practice, even though the predicate covering the two items (or pairs, if it's a relation) *is* disjunctive. The use of an algebraic function in, say, Newtonian mechanics works this way. The fact that the output value of an equation, given its input values *is the same as* the magnitude of an object's force, given its mass and its rate of acceleration, *with respect*, quite generally, to algebra *and*, say, metrics is quite relevant. Something like the respect "algebra and metrics" is a way that things can be the same that matters quite a lot, antecedent to any specific analogical model we may make. This respect of sameness is, then, to be contrasted with the respects underlying Kepler's geometry model – relations between the very specific solids and variously sized spheres just aren't the kinds of relations that we've observed to reflect empirical regularities, so it wouldn't be very relevant if they could be mapped 1-to-1 onto the orbits of the planets, all other things being equal.³⁹ Physical magnitudes have been approximately well behaved so far in terms of being (isomorphic to) the right types

³⁸ This might even provide a way of explicating what such models are, that sameness relations between members of isomorphic structures instantiated by them and their ranges must be stated using disjunctive predicates.

³⁹ So, this is not an a priori argument about geometry and space. On my account, it's perfectly possible that the world might have turned out such that the types of geometric relations that Kepler uses in his model would be relevant to other domains, such as astronomy. If that were what our universe were turning out to be like, then the respects of sameness informing Kepler's model would (have) become entrenched in empirical practice, because the fact that things are the same in those ways would be empirically telling, in ways they are in fact not.

of algebraic functions, so in that case the sameness (and the isomorphic structures of which it's a part) can be legitimately applied to the analysis of analogical arguments.

Another type of legitimate case where the only predicate in common between two compared objects is a disjunctive one is when a model is designed specifically for some purpose. Say, for example, an arrangement of colored blocks on a desk as a model for the layout of houses in a neighborhood. Suppose the model preserves the location-relations between all of the houses on a street by having one block for each house, arranged in an exactly opposite ordering to that of the actual houses. This might entail, then, that house *a* is to the right of house *b*, and so block *a'* is to the left of block *b'*. So, $\langle a, b \rangle$ and $\langle a', b' \rangle$ are both denoted by the disjunctive predicate 'either to the right of or to the left of', which itself might be called a 'right and left orientation-predicate'. In ordinary contexts, the fact that two things are the same with respect to right and left orientation doesn't tend to bear much significance. You can't make many predictions based on such similarities and you can't rely on them to reflect deep regularities. However, the context of a local model is different, specifically because the regularities are part of the design. These sorts of models (unlike many more broadly ranging mathematical models) are specifically designed to maintain and reflect relations that would not otherwise be reliable in this way. They reflect and make regular relations of sameness that are not normally entrenched; in this way, the use of a local model can be looked at as the artificial and specialized route to entrenchment of the respects in which the *relata* of the model and its domain are the same.

There are obvious contexts in which such models are and can be useful, such as in teaching/learning situations where visualization tools are helpful, in the GUI design of digital programs, and in various types of mapping applications. To my mind, however, these kinds of models become most interesting and useful in the creation, criticism and analysis of visual artworks or other types of non-articulate complex symbolic items. In such contexts, in their best, an artificially designed structure-sharing can sometimes take on a life of its own, so that aspects of the model and of its *relata* which had not been part of the initial design come out as revelatory and meaningful. One of my most deeply held philosophical commitments is that works of art contain cognitively meaningful content and that the production of, dissemination of, and engagement with such works constitute centrally important, if often unvalued aspects of inquiry. One way this works in individual artworks is by the arrangement of symbolically meaningful (though not necessarily finitely differentiated) elements within the work in ways that correlate with some structure or complex of structures in the world. Through careful criticism, the working out of these correlations helps reveal insights about the structures being symbolized. This kind of symbolizing can obviously take place at a very low level, such as in propaganda works or in morality tales designed to condition obeisance in children. In its best, however, the artist or designer may start with an idea of what kinds of things and relations are to be symbolized, but, ultimately, internal features of the model chosen irrespective of their symbolic functions take on symbolic meaning within the target domain that may not have been thought out or intended. This is what I mean when I say such items take on a life of their own. It seems clear

enough, for example, that Stanley Kubrick intended to indicate the brutal history of conquest by white men on the North American continent in the haunting – the “shining” – of the Overlook⁴⁰ Hotel. And it seems clear that he intended to indicate, by the *haunting*, that we today still live willfully with the presence of this brutal history in our daily lives, in particular as the haunting takes possession of the white male protagonist, who is at once its victim, because the promise of a leisurely life due to the supposed supremacy of his race and gender has turned out to be an empty ploy, as well as its perpetrator in the present. What isn’t clear, however, is how far the model of these relations within the representational context of the film was intended. For example, when the perpetuation of our criminal past comes to the point of passing on to yet another generation – when Jack, rampaging, chases Danny through the hedge maze with an axe, Danny escapes by retracing his own steps and jumping off the path so that as far as Jack can tell, his footsteps in the snow just end. Transferring this occurrence back onto the target domain, the relations between actors and events modeled seem to suggest a kind of warning that we will continue to be haunted by the murderous theft of land on the continent and the economic and social injustices that are a legacy of slavery unless we are willing and able to retrace at least some of the steps that got us here, to go back and face the history of which we should be terrified and ashamed. As I’ve suggested, filmic/textual cues are widely available within *The Shining* to make it clear that the artist intended to be dealing with the hidden and nearly unspeakable history of the United States and that past’s presence in the present, for all of its concealment. However, whether the

⁴⁰ ‘Overlook’, as in “oversee”, or being an overseer, a person in charge of maintaining the productivity of property, human, animal and terrestrial.

retraced steps were intended to figure into this reading in the proscriptive way I've suggested seems less certain, especially since the action on the part of Danny serves an explicit plot function and intensifies suspense. The point here is that once we start looking at the way a great artwork can symbolize complex systems by alighting on structural similarities to those systems, other aspects of the work become fertile ground for further and directed investigation. Whether or not Stanley Kubrick intended for Danny's retraced steps to suggest that we can outrun our past and prevent it from brutalizing us only if we can step back over the events that led us here, this lesson strikes me as important and right, and it's the kind of thing that drawing focus on structural similarities can help us recognize.⁴¹

E. Conclusion

In the preceding section, I have taken as my starting point the immediate failure of any approach to analogy that attempts to formalize the projection of a predicate onto target cases based on that target's sharing properties with a base analogue, which is denoted by the projected predicate. Any attempt to formalize this necessarily faces the problem that most properties are such that sharing them is irrelevant to the sharing of most other given properties. This worry motivates my approach, according to which, what matters in analogical arguments is an underlying isomorphism between systems containing the items (base and target) of

⁴¹ Especially as I have described this idea using the metaphor of "taking on a life of it's own", Werner Herzog seems to gesture, with different emphasis, at similar possibilities in the following remark: "Your film is like your children. You might want a child with certain qualities, but you are never going to get the exact specification right. The film has a privilege to live its own life and develop its own character. To suppress this is dangerous. It is an approach that works the other way too: sometimes the footage has amazing qualities that you did not expect."

comparison. Something like this seems right, but, on it's own radically underdetermines even a heuristic for dealing with actual applications of analogical arguments. I have proposed my own view, respectivism, to significantly lessen this underdetermination. And to the end of both showing that this lessening needs to be done and that my view has the resources to do it, I have considered two analogical arguments, one good an important in the history of science, and one bad and important in the history of scientific error. And I have shown that, without the emendation of a respectivist criterion, the isomorphism approach can be used to analyze and evaluate good arguments so that they look bad and to analyze bad arguments so that they look good. Finally, I have clarified a critical issue that arises out of my analysis of cases of unilluminating isomorphism (and are jettisoned by my respectivist criterion), namely the status of disjunctive common predicates (and conjunctive respects of sameness). Cases of isomorphism relying on or instantiating sameness relations in such respects are not to be jettisoned simply for that reason; as suggested, some conjunctive respects of sameness become entrenched, either by proving useful in empirical enquiry or by the conventions of a constructed model.

The central direction of my effort here follows in the tradition of Nelson Goodman, his approach to induction and the formulation of his theory of projection. I draw on the notion of "entrenchment" of respects of sameness, an analogue of his entrenchment of predicates in empirical practice. This should not be taken to suggest that I endorse Goodman's particular theory of projection (1983a) or that I actually mean to use that theory to explain or illuminate what I have variously called 'entrenchment'. The details of this theory I leave open here, with just the general

claim that properties become entrenched (and predicates naming them become projectible) by being used with success in past and ongoing empirical practice. The relation to Goodman is much deeper than my loose use of his concept, however. My overall tendency here has been to contrast (very shortly) the shared-properties view as well as the isomorphism view—both of which propose only formal constraints—with an approach that has an explicit place for what might be called “empirical points of contact”. Such points, fundamentally, are necessitated by the fact that induction is a non-formal inference process, and one whose norms shift with the unfolding of inquiry, its successes and failures. The appeal to relative “entrenchment” of respects of sameness that underlie isomorphic structures constitutes a call for empirical points of contact to constrain which formal isomorphic structures we should choose to employ. Without such points, isomorphic structures in keeping with the stated evidence of an analogical argument can be formulated at will which have nothing to do with empirical reality and, if employed, legitimate the drawing of conclusions known antecedently to be false. In this respect, my effort here falls into the tradition not only of Goodman but of Quine as well, as I take a central insight of the latter’s rejection of the analytic-synthetic distinction to be a demand for vigilance in maintaining empirical points of contact with even the most formal-looking sorts of truths about the world.⁴²

⁴² I elaborate on this idea in Chapter 1, section II. C., pp 36–43.

III. Beyond the Monolith

In the preceding section, I dealt with analogies as they are used in arguments, primarily, because this has been the focus of discussion of analogies on the part of philosophers. I should note, however, that throughout this dissertation, I have taken analogies to be a special type of complex similarity relation between objects or systems of objects. Thus, for me (and the theory I have offered), analogies are not arguments, and they are not a class of linguistic items. The use of my theory, therefore, within the context of philosophical concern with analogical arguments should be viewed in this light. Though I don't think that analogical arguments are very important epistemically and I find the focus of philosophers on them as the relevant focus of discussions of analogy to be misguided, I also think that my theory can assist in those discussions in the ways I've indicated.

In general, rather than grounding inductively useful inferences, I think analogies tend to function, within inquiry broadly characterized, more as heuristic devices that help us "see" complex relations in ways that we are unlikely to notice without the help of the analogy, because of the complexity packed into the fact that a given system shares a given structure with another system. This is just the sort of use to which isomorphism is put in mathematics. Thus, beyond the antecedent preoccupations of philosophers with explaining the validity of arguments by analogy, my hope is that my theory can be used to bring to light commonly shared structures in widely disparate domains such as to help develop insights into deep and complex similarities, which are, without the systematic framework my theory

provides, difficult or practically impossible to identify. I believe my concept of structure-predicates can help do this.

While I define the concept of a structure-predicate, in Chapter 4⁴³, I say nothing about such predicates and give no examples of them. One motivation for generating the explication out of this concept is to unify my explication of analogy with my previous explication of similarity, such that analogy ends up being a special kind of similarity more broadly understood. This motivation is fine as far as it goes, as it cooperates with the intuition that analogical relations between objects or sets of objects represent a special type of similarity relation. It also corresponds with the thrust of cognitive psychology's investigations into analogy and similarity, which identify the first as a particular variety of the second. However, beyond the antecedent goal of unifying the account, the notion of structure-predicates holds investigative promise of its own, which I have not yet explored.

In my discussion of analogical arguments, I have mainly treated shared structures as already interpreted; the structures I identify in all of those examples feature actual predicates and relations, not just abstract predicate schemata. This is because we were analyzing already formulated arguments. It bears emphasizing, however, that within the explication, analogical similarity is *abstract* – the isomorphic structures exhibited by two systems are described formally, or in terms of uninterpreted schemata. So, for example, suppose we start looking carefully for any system that might instantiate the following structure:

U: {a, b, c, d, e, f}
F: {a, c, e}

⁴³ P. 141

G: {b, d, f}

R: {⟨a, b⟩, ⟨b, c⟩, ⟨c, d⟩, ⟨d, e⟩, ⟨e, f⟩}

S: {⟨a, c⟩, ⟨b, d⟩, ⟨c, e⟩, ⟨d, f⟩}

It should become immediately evident that, were we able to easily observe structures of this sort (at this level of description), we would start to see them showing up in all sorts of surprising places.⁴⁴ And within the context of my theory, structure-predicates are the terms that denote structures of this sort. In this way, applying the analysis about entrenchment of respects of sameness above to analogically shared structures more abstractly, the hope would be that certain abstract structures come to the fore as interestingly shared by disparate systems – or, certain structure-predicates become entrenched in and through comparative analysis.

The structure I identify above has certain salient features, as are exhibited by the fact that it instantiates common arithmetic features in a set of six numbers. For this reason, it might be a good candidate as a structure to go looking for in some unusual places and possibly interesting places. To this effect, I can offer little more than a hopeful sketch of what an inquiry like this might look like. As I have said, part of the problem with looking for abstract structures in the world is that they are difficult to notice, even with the heuristic that my explication provides. However, much of the infrastructure for doing so already exists on the World Wide Web.

⁴⁴ The set {1, 2, 3, ...6} exhibits this structure in various way, for example, when 'F' and 'G' are 'odd' and 'even' and 'R' is 'one less than' and 'S' is 'two less than'. The structure is also exhibited, however, by a group of three Democrats and three Republicans, each one of whom is slightly smarter than exactly one other, with the exception of the least smart among them, and of whom all give varying amounts of money to his or her respective party, with the least smart of all of them giving the most money.

The most direct way I imagine this proceeding is by editing the sort of information that is relevant for detecting these structures into an already-well-developed database of information, such as Wikipedia. So, along with the usual identifying information under an entry, such as dates, etymology, originators, etc., an entry would list a set of structures known to be exemplified by the entry, at some level of description. Since Wikipedia can be edited by a selected group of its users, I imagine the addition of parameters like this to be able to take on a life of its own and, through development and use, ground the entrenchment of certain structure-predicates such that it could be meaningful and relevant to point out that a system exhibits such-and-such structure. Against claims of this sort would be the background that certain particular, well-known systems exhibit the same structure.

This kind of tool really provides a shorthand, manageable way of dealing with (shared) structural characteristics of empirical phenomena that are too complex to receive ready attention and appreciation. Were a device of this sort to gain wide recognition within a specific database, like Wikipedia, it is easy to imagine adding the same information to all sorts of Web content, for example, as a new classification of meta-tags added to Web pages. Were we able to quickly and easily gain access to the vast variety of structures that every variety of worldly thing exhibits, it seems likely that we would thereby begin to see both some unexpected and possibly illuminating commonalities as well as to appreciate the structural characteristics of empirical phenomena in their own right.

This wide-ranging program falls out of one of the central consequences of my explication of analogy and what I take to be one of its central insights – the radical

overdetermination of analogical similarity. We can find analogies anywhere we look for them, but their overabundance doesn't thereby make them easy to see.

Moreover, the difficulty of seeing them makes it similarly difficult to appreciate how deeply illuminating some of them might be. A program of augmenting and organizing vast stores of information so that we can detect, appreciate, and ultimately refine our understanding of the ubiquitously shared structures in the world suggests a way forward in expanding our insights and abilities in this respect. Who knows what we might find, how we might learn to see what has been in front of us all along?

BIBLIOGRAPHY

- Barker-Plummer, D., J. Barwise and J. Etchemndy 2011. *Language, Proof, and Logic*. CSLI Publications.
- Bureau of International Weights and Measures (BIPM).
http://www.bipm.org/en/si/si_brochure/chapter2/2-1/second.html.
- Carnap, R. 1956. *Meaning and Necessity*, 2nd edn. Chicago: University of Chicago Press.
- ____ 1962. *Logical Foundations of Probability*, 2nd edn. Chicago: University of Chicago Press.
- ____ 1963. "Replies and systematic expositions", *The Philosophy of Rudolph Carnap*, P. A. Schilpp, ed. Evanston, IL: Northwestern University Press, 933–940.
- Chomsky, N. 1989. *Necessary Illusions: Thought Control in Democratic Societies*. Cambridge, Mass: South End Press.
- David, M. 1996. "Analyticity, Carnap, Quine, and Truth", *Noûs*, Vol. 30, Supplement: Philosophical Perspectives, 10, Metaphysics, 281-296.
- DeMorgan, A. 1847. *Formal Logic: or, The Calculus of Inference, Necessary and Probable*.
- Ebbs, G. 1997. *Rule-Following and Realism*. Cambridge, Mass: Harvard University Press.
- Falk, D. 2009. *In Search of Time: Journeys Along a Curious Dimension*. London: National Maritime Museum.
- Frege, G. 1980. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, J. L. Austin, trans. Rev 2nd edn. Evanston: Northwestern University Press.
- Gentner, D. and D. R. Gentner 1983. "Flowing Waters or Teeming Crowds: Mental Models of Electricity", In *Mental Models* (pp. 99-129), D. Gentner and A. L. Stevens, eds.. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Genter, D. and Arthur B. Markman 1995. "Similarity is Like Analogy: Structural Alignment in Comparison", in *Similarity in Language, Thought and Perception*, Cristina Cacciari, ed. Turnhout: Brepols, 111-145.

- Gentner, D and C. Toupin 1986. "Systematicity and surface similarity in the development of analogy". *Cognitive Science*, 10, 277-300.
- Goldstone, R. L. and D. L. Medin. 1994. "Similarity, interactive activation, and mapping: An overview" in *Analogical Connections: Advances in Connectionist and Neural Computation Theory, Vol. 2*, K. J. Holyoak & J. A. Barnden Eds., 321-362.
- Goodman, N. 1961. "About", *Mind* 70, 1-24.
- ____ 1972. "Seven Strictures on Similarity", in *Problems and Projects*. New York: Bobbs-Merrill, 437-447.
- ____ 1983a. "Prospects for a theory of projection" in *The Concept of Evidence*, Peter Achinstein, ed. Oxford: Oxford University Press.
- ____ 1983b. "The New Riddle of Induction", in *Fact, Fiction, and Forecast*, 4th edn. Cambridge, Mass: Harvard University Press, 59-83.
- Holyoak, K. J. and P. Thagard 1989. "Analogical mapping by constraint satisfaction". *Cognitive Science* 13, 295--355.
- Hummel, J. E. and K. J. Holyoak 1997. "Distributed representations of structure: A theory of analogical access and mapping." *Psychological Review*, 104, 427-466.
- James, W. 1892. *Psychology: The briefer course*, Notre Dame: University of Notre Dame Press, 1985.
- Keane, M.T. 1995. "On order effects in analogical mapping: Predicting human error using IAM", in *Seventeenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Kepler, J. 1596. *Mysterium Cosmographicum*.
- ____ 1609. *Astronomia Nova*.
- Lang, K. *Astrophysical Formulae, Volume I*. Boston: Springer.
- Maher, P. 2004. "Probability captures the logic of scientific confirmation", in *Contemporary Debates in Philosophy of Science*, C. R. Hitchcock, ed., 69-93. Blackwell, Oxford, 2004.
- Maher, P. 2007. "Explication Defended", *Studia Logica: An International Journal for Symbolic Logic*, Vol. 86, No. 2, 331-341.

- Markman, A. B. and Deidre Gentner 1990. "Analogical mapping during similarity judgments", in *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, Hillsdale: Erlbaum.
- Medin, D. and Robert L. Goldstone. 1995 "The Predicates of Similarity", in *Similarity in Language, Thought and Perception*, Cristina Cacciari, ed., pp. 83-111.
- Medin, D., Robert L. Goldstone, and Dedre Genter. 1993 "Respects for Similarity." *Psychological Review*. Vol. 100, No. 2, 254-278.
- Moore, G. E. 1942. "Reply to My Critics", in *The Philosophy of G. E. Moore*, P. Schilpp, ed. Evanston, Illinois: Open Court, 663.
- Moore, G. H. 1988. "The emergence of first order logic", in *History and Philosophy of Modern Mathematics*, Aspray, W., Kitcher, P., eds. U. of Minnesota Press, 95-135..
- Müller, A. 2012. "Putnam versus Quine on revisability and the analytic-synthetic distinction", in *Reading Putnam*, Maria Baghramian, M., ed., Oxford: Routledge, 145-178.
- Nelson, R. et al. 2001. "The leap second: its history and possible future", *Metrologia* 38, 509-529.
- Nosofsky, R. M. 1992. "Exemplar-based approach to relating categorization, identification, and recognition", in *Multidimensional models of perception and cognition*, F. G. Ashby ed. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, 363-393.
- Putnam, H. 1975. "The Analytic and the Synthetic" in *Philosophical Papers, Volume 2: Mind, Language and Reality*. Cambridge: Cambridge University Press, 33-69.
- Quine, W.V. 1960. *Word and Object*. Cambridge, Mass: MIT Press.
- ____ 1961. "Two Dogmas of Empiricism", in W. V. Quine, *From a Logical Point of View*, rev. 2nd edn. Cambridge, Mass: Harvard University Press, 20-46.
- ____ 1986. *Philosophy of Logic*, 2nd edn. Cambridge Mass: Harvard University Press.
- Ricardo, D. 1951. *Principles of Political Economy and Taxation*, P. Sraffa ed. Cambridge: Cambridge University Press.

- Russell, B. 1961. "Aristotle's Logic" in *Basic Writings 1903-1959*. London: George Allen & Unwin, 251–258.
- Sen, A. 1988. *On Ethics and Economics*. Oxford: Blackwell Publishing.
- Shepard, R. N. 1962. "The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I", *Psychometrika* 27, 125-140.
- Soames, S. 2003. *Philosophical Analysis in the Twentieth Century, Volume 1: The Dawn of Analysis*. Princeton, NJ: Princeton University Press.
- Sosa, E. 1983. "Classical Analysis," *Journal of Philosophy* 53, 695–710.
- Strawson, P.F. 1963. "Carnap's views on constructed systems versus natural languages in analytic philosophy", *The Philosophy of Rudolph Carnap*, P. A. Schilpp, ed. Evanston, IL: Northwestern University Press, 503–518.
- Tversky, A. 1977. "Features of similarity", *Psychological Review* 84, 327-352.
- Weitzenfeld, J. 1984. "Valid Reasoning by Analogy", *Philosophy of Science* 51: 137-49.
- Wilson, W. H., G. S. Halford, B. Gray, and S. Phillips 2001. "The STAR-2 Model for Mapping Hierarchically Structured Analogs", in *The analogical mind: Perspectives from cognitive science*, D. Gentner, K.J. Holyoak, and B. Kokinov, eds., 125-159. Cambridge, MA: MIT Press.
- Wisniewski, E. J. and D. L. Medin 1990 . "Harpoons and long sticks: The interaction of theory and similarity in rule induction", in *Computational Approaches to Concept Formation*, D. Fisher and M. Pazzani, eds., San Mateo: Morgan Kaufman.